

AD-A252 516



1992

~~THESIS~~/DISSERTATION

A Theoretical Study of the Mechanism of the Alkylation
of Guanine by N-Nitroso Compounds

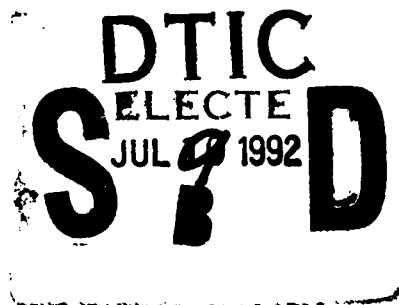
Joel D. Cain, Major

AFIT Student Attending: University of North Carolina

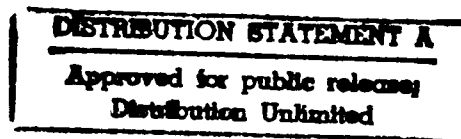
AFIT/CI/CIA- 92-006D

AFIT/CI
Wright-Patterson AFB OH 45433-6583

Approved for Public Release IAW 190-1
Distributed Unlimited
ERNEST A. HAYGOOD, Captain, USAF
Executive Officer



92 7 09 014



92-18005



A THEORETICAL STUDY OF THE MECHANISM OF THE
ALKYLATION OF GUANINE BY N-NITROSO COMPOUNDS

by

Joel D. Cain

A Dissertation submitted to the faculty of the
University of North Carolina at Chapel Hill
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in the Department of Chemistry

Chapel Hill

1992

Approved by:

<u>L. Pedersen</u>	Advisor
<u>Robert S. Pan</u>	Reader
<u>Robert L. King</u>	Reader

JOEL D. CAIN. A Theoretical Study of the Mechanism of the Alkylation of Guanine by N-Nitroso Compounds (Under the direction of Professor Lee G. Pedersen)

ABSTRACT

N-nitroso compounds are potent, organ-specific carcinogens which effect chemical mutations via alkylation of the DNA base guanine. The resulting G:C \rightarrow A:T transition is believed to be due to anomalous base pairing of O⁶-alkylguanine with thymine during replication. The ultimate metabolite involved in the alkylation reaction has generally been thought to be an alkyldiazonium ion or, its decomposition product, a carbocation. In this study, semiempirical (MOPAC) analysis of the enthalpy changes associated with the alkylation on guanine of the O⁶ oxygen, the purported promutagenic site, and the N⁷ nitrogen by alkyldiazonium ions and by carbocations indicate that the alkyldiazonium ion is the more likely ultimate mutagen. However, the deprotonation of the N¹ nitrogen, as observed in x-ray studies of O⁶-methylguanine, was not apparent in these semiempirical calculations. Subsequent calculations on the possible involvement of water in the loss of the N¹ hydrogen were performed using both semiempirical and density-functional (DGauss) techniques. The density-functional calculations proved comparable to high-level traditional *ab initio* calculations on model reactions and allowed such rigor to be reasonably applied to a guanine-sized system. A two-step mechanism is proposed in which an intact alkyldiazonium ion attacks the O⁶ position and, then, deprotonation at N¹ occurs with water acting as a proton acceptor.

For my wife, Sheila, whom I love dearly



Accession For	
NTIS GRA&I	<input checked="" type="checkbox"/>
DTIC TAB	<input type="checkbox"/>
Unannounced	<input type="checkbox"/>
Justification	
By	
Distribution/	
Availability Codes	
Dist	Avail and/or Special
A-1	

ACKNOWLEDGEMENTS

I would like to express my deep appreciation to my research advisor, Dr. Lee Pedersen, for his unfailing patience and support during all of my graduate work at Carolina. He has continued to be a constant source of advice and encouragement. I would also like to thank the other members of my committee, Dr. Robert Parr, Dr. Charles Johnson, Dr. Richard Hiskey, and Dr. Nancy Thompson. Over several years, they have been a continuing inspiration.

There are several people who have also been instrumental in the success of this project. I would like to thank Dr. David Deerfield for his advice and patient explanations when I first started in computational chemistry, Dr. Thomas Darden and Dr. Howard Smith for their helpful assistance in using the computers at NIEHS, and Dr. Lee Bartolotti for his help with UniChem and DGauss. I would also like to thank Warren Pope and NIEHS for the use of their facilities and computer equipment.

For the opportunity to do this work, I would like to thank the United States Air Force Academy and the Air Force Institute of Technology for their sponsorship of my program.

Last, and most of all, I would like to thank my parents, family, and, especially, my wife, Sheila, for their unending support and love. Without them, this project could not have been a success.

TABLE OF CONTENTS

	Page
LIST OF TABLES	vii
LIST OF FIGURES	ix
LIST OF ABBREVIATIONS	xi
 Chapter	
I. Introduction and Background	1
1.1 Introduction	1
1.2 Biochemistry of N-Nitroso Compounds	3
1.3 Mutagenicity of N-Nitroso Compounds	10
References	16
II. Computational Methods	21
2.1 Transition State Optimization	21
2.2 Semiempirical Calculations with MOPAC	23
2.3 MOPAC 5.0 Optimization Methodology	26
2.4 Eigenvector Following in MOPAC 6.0	30
2.5 Density-Functional Theory	33
References	40
III. Preliminary Semiempirical Calculations	43
3.1 Introduction	43
3.2 Alkyldiazonium Ions	44
3.3 Guanine	50
3.4 Methyldiazonium Ion - Formamide Transition State	56
3.5 Typical S _N 2 Transition States	62
3.6 Conclusions from Preliminary Calculations	66
References	68

IV. Semiempirical Calculations on Guanine Transition States	71
4.1 Introduction	71
4.2 Overall Shape of the Potential Surface (MOPAC 5.0)	72
4.3 Transition States (Bartels's Method - MOPAC 5.0)	77
4.4 Eigenvector Following Calculations (MOPAC 6.0)	83
4.5 Conclusions	88
References	91
V. The Deprotonation of the N ¹ Nitrogen	94
5.1 Introduction and Methods	94
5.2 The Effect of Water on Formamide-Based Models	97
5.3 The Effect of Water on Guanine-Based Systems	107
5.4 Forced Deprotonation of the N ¹ Nitrogen	123
5.5 Conclusions	125
References	131
Appendix	133

LIST OF TABLES

Table 3.1:	Geometry of the Methyldiazonium Ion	44
Table 3.2:	Diazonium Ion Enthalpic Calculations	47
Table 3.3:	<i>Ab Initio</i> Dissociation Energies of Alkyldiazonium Ions	49
Table 3.4:	Bond Lengths in 9-Methylguanine	50
Table 3.5:	9-Methyl-O ⁶ -Methylguanine	53
Table 3.6:	Hydrogen-Bonding Distances in G:C Pair	54
Table 3.7:	Selected Results of Formamide Calculations	58
Table 3.8:	Selected Results of Calculations on O ⁶ -Methylated Formamide (with +1 Charge)	58
Table 3.9:	Transition State Calculations for the O ⁶ Methylation of Formamide by the Methyldiazonium Ion	60
Table 3.10:	Intrinsic Barriers for Degenerate S _N 2 Reactions	66
Table 4.1:	PM3 Heat of Formation Results of Coordinate-Driving Calculations	73
Table 4.2:	PM3 Geometries of the Pre-Transition State Ion-Molecule Complexes at the O ⁶ of Guanine	75
Table 4.3:	PM3 Geometries of the Pre-Transition State Ion-Molecule Complexes at the N ⁷ of Guanine	75
Table 4.4:	PM3 Geometries of the Alkylated Products	76
Table 4.5:	PM3 Geometries of the Transition State (Bartels's Method) for Alkylation at the O ⁶ of Guanine	78

Table 4.6:	PM3 Geometries of the Transition State (Bartels's Method) for Alkylation at the N ⁷ of Guanine	78
Table 4.7:	PM3 Heats of Reaction, Activation Enthalpies and Intrinsic Barriers	80
Table 4.8:	PM3 Force Constants for Transition States Optimized by Bartels's Method	81
Table 4.9:	PM3 Force Constants for Some Typical S _N 2 Transition States Optimized by Bartels's Method	82
Table 4.10:	PM3 Geometries of the Alkylated Products (Eigenvector Following)	84
Table 4.11:	PM3 Geometries of the Transition State (Eigenvector Following) for Alkylation at the O ⁶ of Guanine	85
Table 4.12:	PM3 Geometries of the Transition State (Eigenvector Following) for Alkylation at the N ⁷ of Guanine	85
Table 4.13:	PM3 Heats of Reaction, Activation Enthalpies and Intrinsic Barriers - Eigenvector Following	86
Table 4.14:	PM3 Force Constants for Transition States Optimized by Eigenvector Following	87
Table 5.1:	Bond Lengths in Guanine	111
Table 5.2:	Bond Lengths in Deprotonated (N ¹) Distal O ⁶ -Methylguanine	111
Table 5.3:	ESP (MOPAC 6.0) Calculated Charges in Guanine, O ⁶ -Methylguanine, and the MDO6G Transition State (with N ¹ Hydrogen)	116

LIST OF FIGURES

Figure 1.1: Representative N-Nitroso Compounds	2
Figure 1.2: Standard Numbering in DNA Bases	6
Figure 1.3: Base Pairing of O ⁶ -Alkylguanine with Thymine <i>versus</i> Normal G:C Pairing	13
Figure 1.4: The G:C --> A:T Transition	14
Figure 3.1: Methyldiazonium Ion	45
Figure 3.2: Distal and Proximal Forms of 9-Methyl-O ⁶ -Methylguanine	52
Figure 3.3: Hydrogen Bonding in G:C Base Pair	55
Figure 3.4: Methylation of Formamide by Methyldiazonium Ion	57
Figure 3.5: Plots of Geometry <i>versus</i> $\Delta_{\text{rxn}}H$ for $X^- + \text{CH}_3\text{F} \longrightarrow \text{CH}_3\text{X} + \text{F}^-$	64
Figure 3.6: Plots of Geometry <i>versus</i> $\Delta_{\text{rxn}}H$ for $X^- + \text{CH}_3\text{OH} \longrightarrow \text{CH}_3\text{X} + \text{OH}^-$	65
Figure 5.1: Optimized Geometries of Formamide-Based Systems .	98
Figure 5.2: The Effect of Water on Formamide and O-methyl Formamide (+1)	101
Figure 5.3: The Effect of Water on the Transition State for the Methylation of Formamide by the Methyldiazonium Ion	102
Figure 5.4: Calculated Atomic Charges in Formamide-Based Systems	104

Figure 5.5: Reaction Profile for the Methylation of Formamide by the Methyldiazonium Ion	106
Figure 5.6: Selected Optimized Geometry Parameters for Guanine-Based Systems	109
Figure 5.7: Selected Charges in Guanine-Based Systems	110
Figure 5.8: The Effect of Water on Guanine and O ⁶ -Methylguanine (+1)	113
Figure 5.9: The Effect of Water on the MDO6G Transition State (+1) - Results of Transition State Optimization	114
Figure 5.10: DGauss Electron Density Diagram for Guanine	117
Figure 5.11: DGauss Electron Density Diagram for O ⁶ -Methylguanine (+1)	118
Figure 5.12: DGauss Electron Density Diagram for the MDO6G Transition State (+1)	120
Figure 5.13: Reaction Profile for the O ⁶ Methylation of Guanine by the Methyldiazonium Ion	122
Figure 5.14: Heat of Formation <i>versus</i> N ¹ -H Distance for Forced Deprotonation at N ¹	124
Figure 5.15: MOPAC Reaction Profile for the Loss of the N ¹ Hydrogen to a Water Molecule	126

LIST OF ABBREVIATIONS

A	- adenine
AM1	- Austin model 1
BFGS	- Broyden-Fletcher-Goldfarb-Shanno
C	- cytosine
CI	- configuration interaction
DES	- diethylsulfate
DFT	- density-functional theory
DMN	- N,N-dimethylnitrosoamine
DNA	- deoxyribonucleic acid
EMS	- ethyl methanesulfate
ENNG	- N-ethyl-N'-nitro-N-nitrosoguanidine
ENU	- N-ethyl-N-nitrosourea
G	- guanine
MGMT	- O ⁶ -methylguanine-DNA methyltransferase
MNDO	- modified neglect of diatomic overlap
MNNG	- N-methyl-N'-nitro-N-nitrosoguanidine
MNU	- N-methyl-N-nitrosourea
MNUureth	- N-methyl-N-nitrosourethane
MOPAC	- Molecular Orbital Package
MP	- Møller-Plesset
PM3	- parametric method 3
SCF	- self-consistent field
T	- thymine

CHAPTER 1

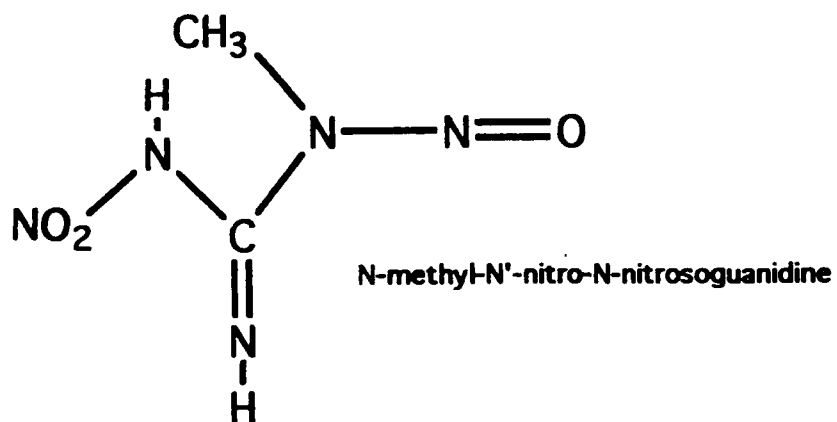
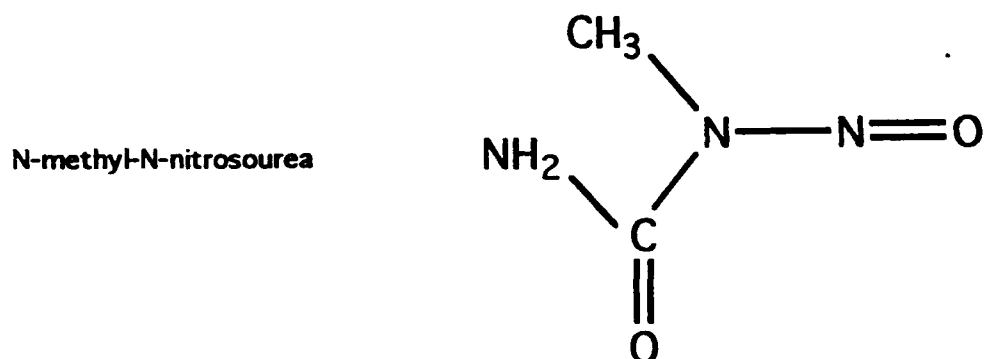
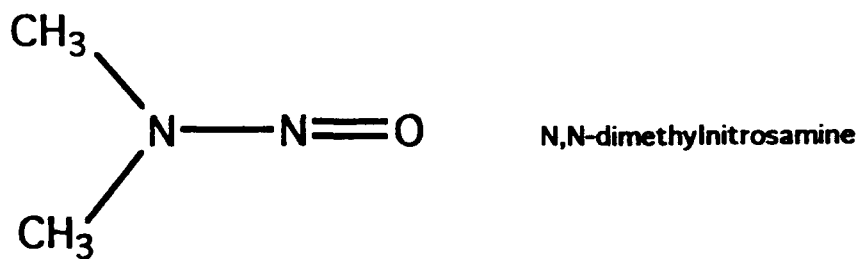
Introduction and Background

1.1 Introduction

Mutagenesis induced by various chemical substances has been a subject of study for some time. Since Auerbach and Robson established mustard gas as a potent chemical mutagen in 1946 [1], the list of substances identified or implicated as mutagenic and/or carcinogenic has grown tremendously [2-4]. One important class of mutagens are those which act by alkylation of various sites in nucleic acids [5]. Among these are the N-nitroso compounds such as the nitrosamines, the nitrosoureas, the nitrosoguanidines, and the nitrosourethanes [3,6] (see Figure 1.1). Considerable experimental and theoretical work has been done to help understand more about these alkylating mutagens in terms of their metabolism to ultimate reactive species, the mechanism of their attack on nucleic acids, and how the resulting adducts lead to mutations and cancer [3-9].

The purpose of this study is to computationally characterize the transition states involved in the alkylation of guanine bases in nucleic acids by certain possible metabolites of N-nitroso compounds. The alkylation mechanism will be studied using semiempirical methods and density-functional theory (DFT). These are the only really practical computational methods to study the bond breaking and bond forming occurring in these reactions since

Figure 1.1 - Representative N-Nitroso Compounds



traditional *ab initio* calculations on guanine-sized systems at any reasonably high level of theory would be computationally prohibitive. The relationships among the energies and geometries of the optimized structures will be discussed, and an attempt will be made to infer something about the reaction mechanism. The goal is to identify the ultimate reactive metabolite of N-nitroso compounds which attacks certain nucleophilic sites in guanine to effect alkylation and to help develop a theoretical model of how the alkylation occurs.

The remainder of this chapter outlines some essential aspects of the biochemistry of N-nitroso compounds and reviews various experimental observations and theoretical results in the history of N-nitroso compounds as identified chemical mutagens. In the next chapter, the computational techniques used in this study are described. In succeeding chapters, the results of certain preliminary calculations are reported to help lend credibility to the chosen semiempirical methodology, the results of the semiempirical calculations on the guanine transition states are discussed, and the application of DFT techniques are presented.

1.2 Biochemistry of N-Nitroso Compounds

The carcinogenicity of N-nitroso compounds was first demonstrated in 1956 by Magee and Barnes who reported malignant liver tumors in rats after administration of dimethylnitrosamine [10]. These compounds have since been studied extensively, both by experimentalists and theoreticians. This emphasis on N-nitroso compounds is easily understood. They are potent, organ-specific

carcinogens in every species tested so far. Also, N-nitroso compounds are prevalent in our diets (e.g., nitrite-cured meats and alcoholic beverages), and in our environment (e.g., tobacco smoke and various industries), and they are even formed inside the human body from ingested non-carcinogenic precursors [11,12].

It has been shown that N-nitroso compounds are rapidly and uniformly distributed throughout the body after injection into rats and mice [3, 13,14]. However, for the nitrosamines, reactions with macromolecules are evident only in tissues containing the cytochrome P450-dependent mixed function oxidases which serve to hydroxylate foreign substances in the body to make them more water soluble and facilitate excretion [3,15]. A mechanism which has been proposed for the promutagenic metabolism of N,N-dimethylnitrosamine (DMN) involves the oxidase enzyme hydroxylating a methyl group on DMN which is then rapidly lost as formaldehyde leaving a diazohydroxide which decomposes to form a diazonium ion.

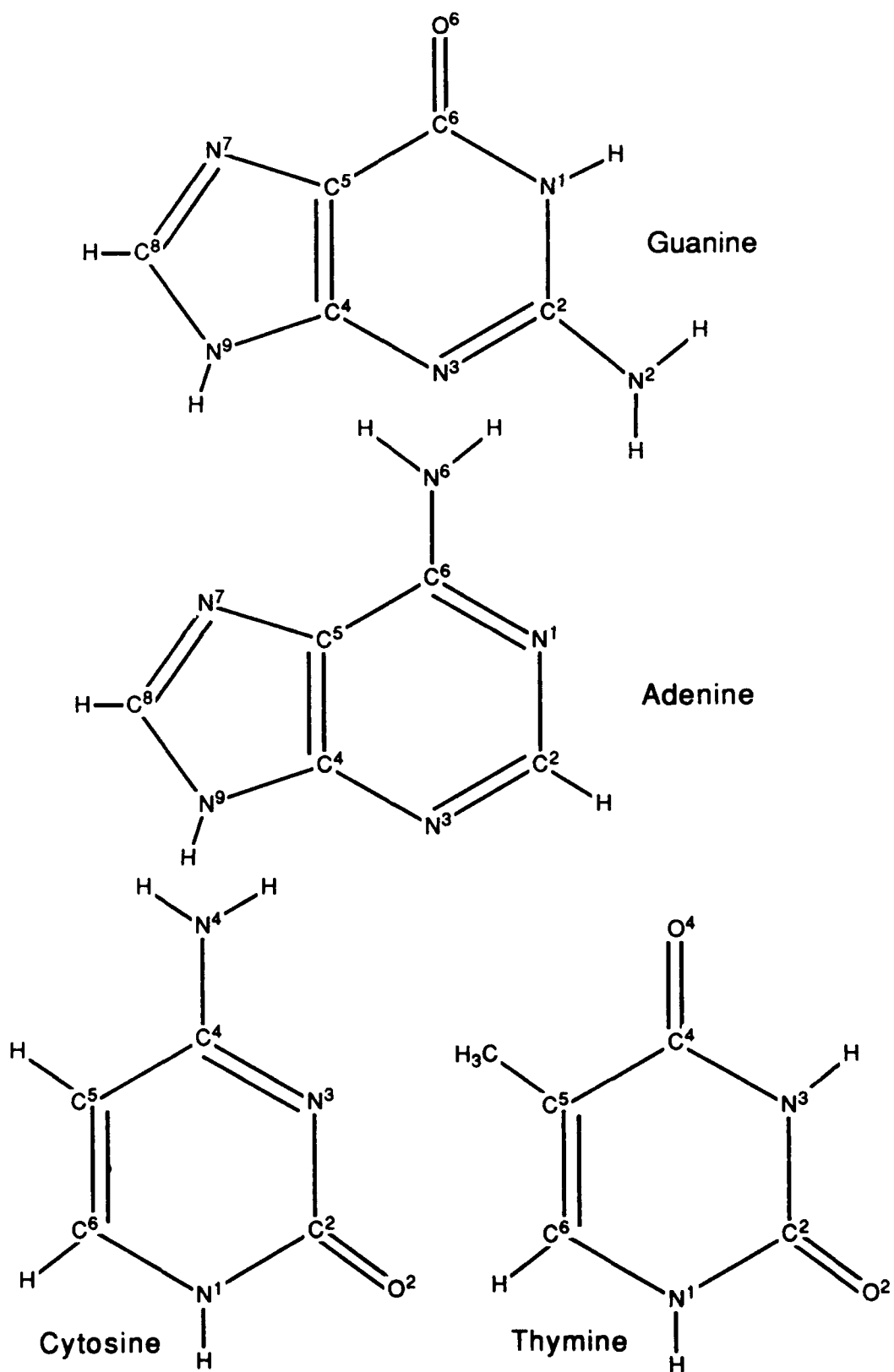
For nitrosamides, such as N-methyl-N-nitrosourea (MNU), N-methyl-N'-nitro-N-nitrosoguanidine (MNNG), and N-methyl-N-nitrosourethane (MNUreth), the enzymatic activation is apparently not necessary since nitrosamides react similarly with macromolecules in all tissues [3,16-18]. The metabolism of nitrosamides is thought to involve spontaneous multi-step decomposition to form, ultimately, the same diazohydroxide and diazonium ion formed as a metabolite of DMN [3]. There is an increased reactivity of MNNG and MNUreth in tissues which contain greater proportions of sulfhydryl groups (e.g., cysteine) which

catalyze the breakdown of these compounds; however, there is no similar effect for MNU [3,18-22].

The rate of metabolism of nitrosamines can be measured through the use of ^{14}C labeling and subsequent monitoring of radioactivity in exhaled CO_2 formed as a by-product of nitrosamine metabolism [3]. Nitrosamide decomposition can be monitored through radioactive counting of tissue samples after administering radioactively-labeled nitrosamides [23]. Nitrosamide decomposition occurs more rapidly than nitrosamine metabolism. Swann and Magee administered MNU and N-ethyl-N-nitrosourea (ENU) to rats at 100mg/kg of body weight for MNU and 200 mg/kg for ENU and found half-lives of two minutes and five minutes respectively [17,24]. On the other hand, metabolism of DMN has been found to take three to six hours depending on the method of administration [3,25]. Still, these are short times when compared to the carcinogenic time table.

Several studies have been aimed at understanding how these chemical agents alkylate DNA, but, as yet, the precise mechanism is unknown. What is known is that the result is a DNA-mutagen adduct with an alkyl group becoming covalently bound to some nucleophilic site on the DNA. There are many such nucleophilic sites. Through the use of radioactively-labeled nitrosoureas, Singer *et. al.* found that about 25% of the alkylation caused by MNU was on the DNA phosphate backbone while, for ENU, phosphate alkylation represented about 65% of the total alkyl groups bound to the DNA [26]. Singer [27] found that for ENU and N-ethyl-N'-nitro-N-nitrosoguanidine (ENNG), over 80% of the alkylation occurs at oxygen centers with, other than phosphate and ribose oxygens, the most likely oxygen

Figure 1.2 - Standard Numbering in DNA Bases



sites to be alkylate in double-stranded DNA being the O² position of thymine and the O⁶ position of guanine (see Figure 1.2).

In general, methylating agents are more reactive with nucleic acids than are ethylating agents which are, in turn, more reactive than propylating agents [6]. The effects of strandedness on reactivity of sites in DNA is apparent for the N¹ position of adenine and the N³ position of cytosine which, presumably due to being involved in hydrogen bonding, are much less reactive when in double-stranded DNA [6,28-30], but they are significantly alkylated even in double-stranded forms [6,31,32]. However, the reactivities of oxygen centers are unaffected by strandedness since even those involved in hydrogen bonding are vulnerable through their unbonded electron pairs [6]. Even with all this, the actual extent of DNA modification is low. In most experiments, 0.1% or less of the nucleic acid bases are affected [3].

Support for the idea that alkylation by N-nitroso compounds occurs via common intermediate types comes from observations of similar alkylation patterns for analogous (methyl, ethyl, etc.) compounds [3,33,34]. At one time, the diazoalkanes, known to be derivatives of nitrosamides, were thought to be involved [3,35]. However, this proved to be inconsistent with a subsequent isotope ratio study using MNU. In that study [36], Lawley and Shah used MNU which had been methyl-labeled with ³H and ¹⁴C. Treated DNA was subsequently hydrolyzed and separated by column chromatography to yield products of methylation at several sites on nucleic acid bases. Comparison of radioactivity in products from alkylation with ¹⁴CH₃-, C³H₃-, and ¹⁴C³H₃-labeled MNU showed that the ³H/¹⁴C

ratio in the products was the same as that in the reactants indicating intact transfer of the methyl group. Another study by Sussmuth *et. al.* [37] used trideuterated MNNG to treat DNA. Subsequent analysis by mass spectrometry and, then, by nuclear magnetic resonance again showed intact transfer of the methyl group. This type of work has helped lead to the type of mechanism described previously where the N-nitroso compounds were shown producing a diazonium ion.

The prevalent question concerning the identity of the ultimate reactive species has been whether the alkyldiazonium ion directly attacks the nucleic acid base, or whether the alkyldiazonium ion first decomposes to form a carbocation which, in turn, reacts with the nucleic acid. In the first case, the reaction would be S_N2 , and, in the latter, S_N1 [3,4]. A theoretical study by Ford and Scribner [38] gave semiempirical reaction enthalpies which showed alkyldiazonium ion decomposition to carbocation and N_2 to be significantly endothermic indicating that the formation of reactive carbocations are not energetically favored. However, a later study by Freceer and Miertus [39] used *ab initio* (4-31G) methods with a solvent model and found a low positive reaction enthalpy for methyldiazonium ion decomposition and an exothermic ethyldiazonium ion decomposition which would facilitate carbocation formation.

However, Ford and Scribner [38] cite Friedman [40] as concluding that there is no experimental basis for carbocations in the reactions of primary alkyldiazonium ions. There is, in fact, experimental evidence that carbocations are not involved in the

alkylation of nucleic acids by certain N-nitroso compounds. Park, *et. al.* [41] found that exposure of rats to N-nitrosodi-n-propylamine led to formation of 7-n-propylguanine and not 7-isopropylguanine. Significant free carbocation would have resulted in rearrangement and detection of the isopropyl adduct. Scribner and Ford [42] treated rats with di-n-propylnitrosamine and detected the formation of 7-n-propylguanine and O⁶-isopropylguanine. They suggest that rather than being evidence for the presence of carbocations (else, significant 7-isopropylguanine would have found) their results are, instead, consistent with a different type of alkylation occurring at the less nucleophilic O⁶ position which may involve a looser transition state which allows rearrangement during the bimolecular reaction. In a subsequent MNDO study [43], Ford and Smith found unexpected bond order relationships at the transition state for the reactions of the methyl- and ethyldiazonium ions with various nucleophilic sites. They found that the bond orders of the bonds being broken and the bonds being formed were correlated positively - the longer the breaking bond, the longer the forming bond. Their findings were the opposite of that expected in terms of the Hammond postulate which would lead to a negative correlation [44] and lends further credibility to the unusual nature of these types of transition states.

Ford and Scribner [38] used semiempirical methods (MNDO) to study the reactions of methyl- and ethyldiazonium ions with various model compounds (representing the oxygen and nitrogen nucleophilic sites in DNA) in an effort to give a theoretical explanation for the observation that ethylating agents react more with oxygen sites (as

compared to nitrogen sites) than do methylating agents [45,46]. Their results showed that methylation of a model nitrogen site was kinetically favored over methylation of a model oxygen center, but ethylation for the two sites were more kinetically equitable [38]. Miertus and Trebaticka [47] reported a similar study of methyl and ethyl carbocations using both *ab initio* and semiempirical (MINDO/3) methods. The results of their calculations of the interactions of the carbocations with various oxygen and nitrogen sites in nucleic acid bases indicated that, for methylation, the N⁷ position of guanine is the thermodynamically preferred site while, for ethylation, the oxygens in guanine and cytosine are, along with the N⁷ of guanine, favored sites.

1.3 Mutagenicity of N-Nitroso Compounds

In early experimental work with agents which alkylate DNA, comparisons of ultraviolet absorption spectra of alkylated bases (obtained from chromatographically-separated products of hydrolysis of mutagen-treated DNA) with spectra of known compounds showed that guanine was the most likely base to be alkylated, and that, on guanine, it was the N⁷ ring nitrogen which was the most reactive site towards alkylation [48-51, reviewed in 2]. In 1969, Loveless and Hampton reported the mutagenicity of both MNU and ENU in a study which indicated extensive N⁷ methylation by MNU but only a trace of N⁷ ethylation by ENU and concluded that N⁷ alkylation was not a critical event in the mutagenic mechanism [52]. Later, Loveless isolated an O⁶-methylated product from treatment with MNU and was the first to suggest a relevance of O⁶ alkylation

to the mutagenic and carcinogenic properties of N-nitroso compounds [53]. He reasoned that the anomalous base pairings believed to be involved in producing mutations could be accounted for by O⁶ alkylation of guanine and subsequent loss of the N¹ hydrogen (which is involved in Watson-Crick hydrogen bonding) [53].

There has been additional support for the significance of O⁶ alkylation in the mutagenic mechanism. In 1974, Goth and Rajewsky [54] reported a correlation between O⁶-ethylguanine elimination and the likelihood of carcinogenesis in certain tissues in rats. ENU demonstrates a carcinogenic specificity for the nervous systems which is not consistent with the lack of enzymatically-controlled metabolism mechanism [3]. Goth and Rajewsky found that the half-life of O⁶-ethylguanine in brain tissue was much longer than in liver tissue (about 220 hours versus about 30 hours) and also much longer than the half-life of N⁷-ethylguanine (about 90 hours). In 1980, Newbold *et. al.* [55] reported positive correlations between the carcinogenicity of certain alkylating agents (including MNU) and their ability to alkylate the O⁶ oxygen of guanine. In 1985, van Zeeland *et. al.* [56] reported a study of the mutations caused by ENU, ENNG, ethyl methanesulfate (EMS), and diethyl sulfate (DES) and found that, although they differed in mutagenic potency with ENNG > ENU > DES > EMS, their mutagenic activities were similar when plotted against amount of O⁶-ethylguanine formed. More recently, in 1989, Rudiger *et. al.* [57] found reduced levels of O⁶-methylguanine-DNA methyltransferase (MGMT), an enzyme involved in the removal of methyl groups from the O⁶ of guanine, in tissues from lung cancer patients. Isowa *et. al.* later found an MGMT

deficiency in human liver tumors [58]. These are just a few of the findings which have led to the general conclusion that alkylation of the O⁶ position of guanine is the critical, promutagenic event for DNA alkylating agents like N-nitroso compounds [59,60-64].

The question of how alkylation at the O⁶ position causes mutations can be rationalized in terms of Loveless's original suggestion that alkylation at O⁶ leads to the loss of the hydrogen at the N¹ position and the resulting modification to hydrogen bonding characteristics. An x-ray study by Parathasarathy and Friley [65] reported the crystallographic geometry of O⁶-methylguanosine. They confirmed the lack of a hydrogen at the N¹ position and an increase in the C⁶-O⁶ bond lengths and a decrease in the C⁶-N¹ bond length (as compared to unalkylated guanine [66]) as would be expected as the C⁶-O⁶ bond becomes more like a single bond and the C⁶-N¹ bond attains double bond character all as a result of the alkylation at O⁶. In light of this, additional theoretical support for the promutagenicity of O⁶ alkylation comes from semiempirical calculations by Duncan and Davies [67] which showed that the loss of the N¹ hydrogen from alkylated guanine occurs more easily when the alkylation is at the O⁶ position as compared to the N⁷ position.

The loss of the N¹ hydrogen could result in anomalous G:T base pairing, as shown in Figure 1.3, which could effect the well-documented G:C → A:T transition, the predominant mutation resulting from alkylating mutagens (see Figure 1.4) [59]. The G:C → A:T transition, in which as O⁶-alkylated guanine act like an adenine during replication by coding for a thymine in the daughter strand, has been shown to account for practically 100% of mutations due

Figure 1.3 - Base Pairing of O⁶-Alkylguanine with Thymine
versus Normal G:C Pairing

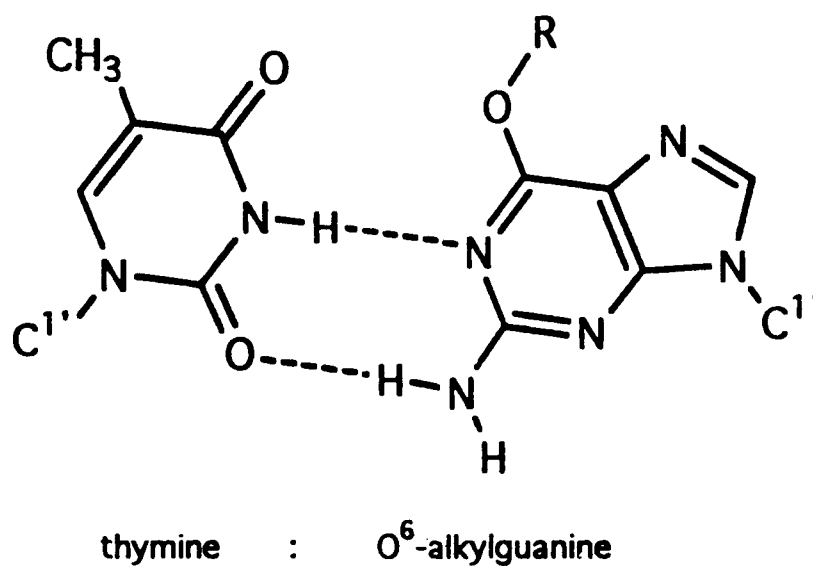
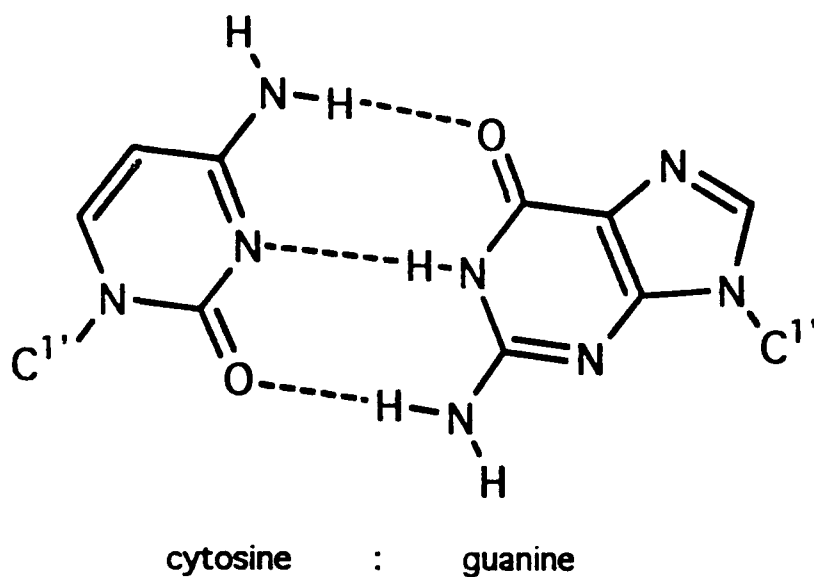
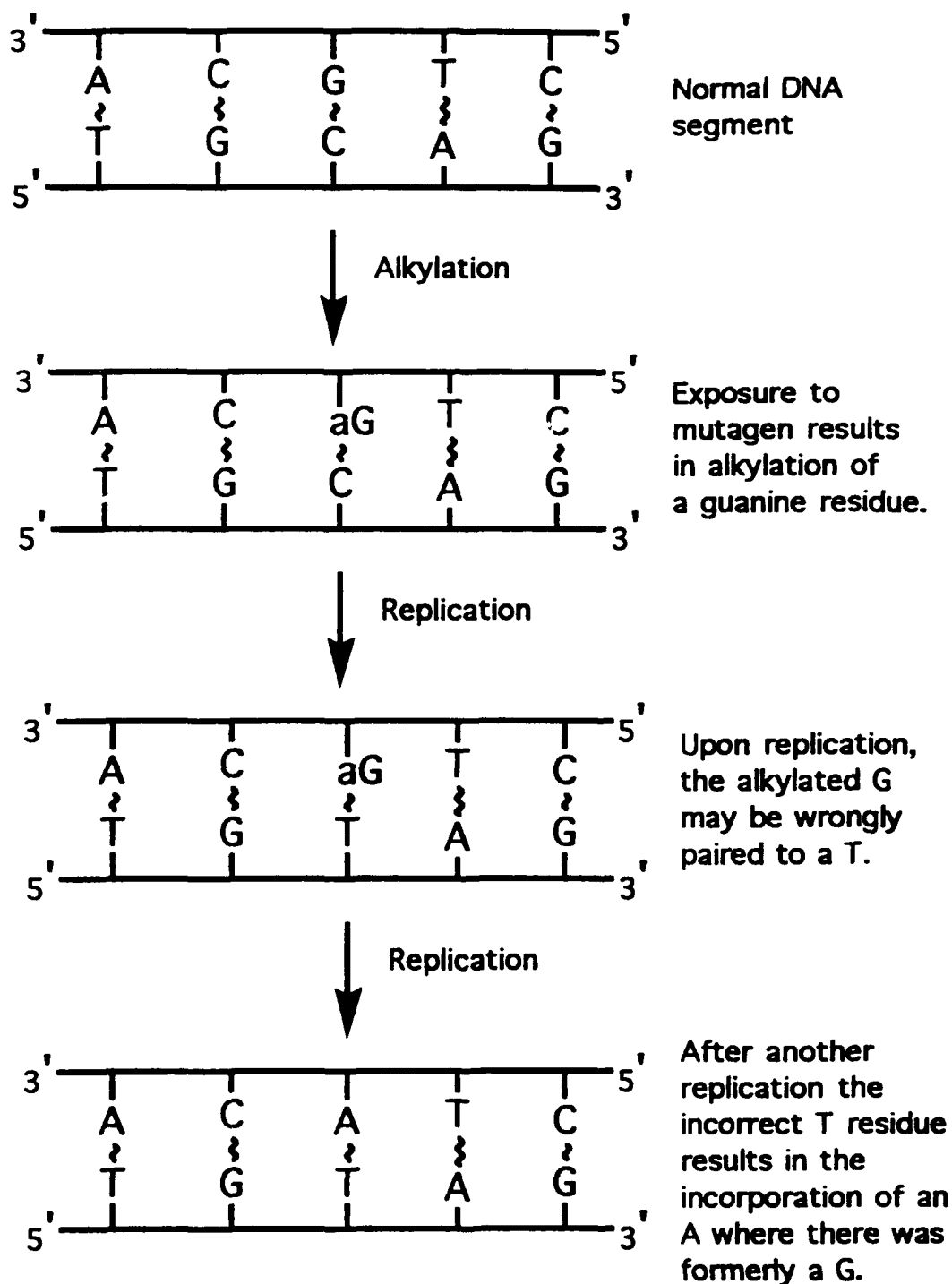


Figure 1.4 - The G:C → A:T Transition



Note: In each replication, only the affected daughter strand and its corresponding new one is shown.

to MNU and 73% of mutations due to ENU [7]. Furthermore, an x-ray diffraction study of a self-complementary double strand containing two O⁶-methylG:T base pairs showed the similarity of the mutated base pair to the normal G:C base pair and suggested that this similarity may contribute to a lack of recognition by repair enzymes [68]. By contrast, Ludlum had earlier shown that the base pairing properties of N⁷-methylguanine and normal guanine are very similar [69].

Most of the theoretical mechanistic studies of mutagenesis by N-nitroso compounds have examined the enthalpies and/or energies of the various reactions believed to be involved in the metabolism and decomposition of these compounds and the reactions of the presumed ultimate reactive species with models of nucleophilic sites in nucleic acids. Less work has been reported in the calculation of the transition states involved. This is probably due, in part, to the difficulty of transition state calculations and to the problem of rigorous computations on nucleic acid-sized molecules. The goal of this study is to perform transition state calculations on these type reactions in order to, first, determine the ultimate reactive metabolite of N-nitroso compounds and, second, to try and understand something about the alkylation-induced deprotonation of the N¹ nitrogen. In the next chapter, the computational methods used in this study are discussed.

References

1. Auerbach, C. and Robson, J.M. (1946) *Nature* , 157, 302.
2. Lawley, P.D. (1989) *Mutation Research* , 213, 3-25.
3. Margison, G.P. and O'Connor, P.J. (1979) in Chemical Carcinogens and DNA. Volume 1, CRC Press, pp.111-159.
4. Singer, B. and Grunberger, D. (1983) Molecular Biology of Mutagens and Carcinogens, New York, Plenum Press.
5. Richardson, F.C. and Richardson, K.K. (1990) *Mutation Research*, 233, 127-138.
6. Singer, B. and Kusmierek, J.T. (1982) *Annual Review of Biochemistry*, 51, 655-693.
7. Singer, B. and Essigmann, J.M. (1991) *Carcinogenesis*, 12, 949-955.
8. Chemical Carcinogens, Second Edition, Vol. 1 and 2, Searle, C.E., ed., ACS Monograph 182, American Chemical Society, Washington, D.C., 1984.
9. Harris, C. (1991) *Cancer Research*, 51, 5023s-5044s.
10. Magee, P.M. and Barnes, J.M. (1956) *British Journal of Cancer*, 10, 114-122.
11. Preussmann, R. and Spiegelhalter, B. (1989) *Interdisciplinary Science Reviews*, 14, 248-253.
12. Tricker, A.R. and Preussmann, R. (1991) *Mutation Research*, 259, 277-289.
13. Magee, P.N. (1956) *Biochemical Journal*, 64, 676.
14. Tjalve, H. and Johansson, E.D. (1977) *Acta Pharmacol. Toxicol, Supp 1*, 41, 150.

15. Weininger, S.J. and Stermitz F.R. (1984) Organic Chemistry, Academic Press, pp. 919-920.
16. Swann, P.F. and Magee, P.N. (1968) *Biochemical Journal*, 110, 39.
17. Swann, P.F. and Magee, P.N. (1971) *Biochemical Journal*, 125, 841-847.
18. Schoental, R. and Rive, D.J. (1965) *Biochemical Journal*, 97, 466.
19. Schulz, U, and McCalla, D.R. (1969) *Canadian Journal of Chemistry*, 47, 2021.
20. Lawley. P.D. and Thatcher, C.J. (1970) *Biochemical Journal*, 116, 693.
21. Wheeler, G.P. and Bowden, B.J. (1972) *Biochemistry and Pharmacology*, 21, 265.
22. Frei, J.V. and Lawley, P.D. (1976) *Chemico-Biological Interactions*, 13, 215.
23. Alarif, A and Epstein S.S. (1975) in N-Nitroso Compounds in the Environment, Bogovski, P., Walker, E.A., and Davis, W., eds., IARC Scientific Publications No. 9, pp. 215-219.
24. Swann, P.F. (1968) *Biochemical Journal*, 110, 49-52.
25. Craddock, V.M. (1971) *Journal of the National Cancer Institute*, 47, 889.
26. Singer, B., Sun, L., and Fraenkel-Conrat, H. (1975) *Proceedings of the National Academy of Sciences*, 72, 2232-2236.
27. Singer, B. (1976) *Nature*, 264, 333-339.
28. Singer, B. (1977) *Journal of Toxicology and Environmental Health*, 2, 1279-1295.

29. Singer, B. (1982) in Molecular and Cellular Mechanisms of Mutagenesis, Lemontt, J.F. and Generoso, W.M., eds., New York, Plenum Press.
30. Jensen, D.E. and Reed, D.J., (1978) *Biochemistry*, 17, 5098-5107.
31. Bodell, W.J. and Singer, B. (1979) *Biochemistry*, 18, 2860-2863.
32. Beranek, D.T., Weis, C.C. and Swenson, F.H. (1980) *Carcinogenesis*, 1, 595-606.
33. Milligan J.R., Hirani-Hojatti, S., Catz-Biro, L. and Archer, M.C. (1989) *Chemico-Biological Interactions*, 72, 175-189.
34. Elespuru, R.K., Stupar, L.L., and Gordon, J.A. (1991) *Carcinogenesis*, 12, 1161-1167.
35. Magee, P.N. and Barnes J.M. (1967) *Advances in Cancer Research*, 10, 163.
36. Lawley, P.D. and Shah, S.A. (1973) *Chemico-Biological Interactions*, 7, 115-120.
37. Sussmuth, R., Haerlin, R., and Lingens, F. (1972) *Biochimica et Biophysica Acta*, 269, 276-286.
38. Ford, G.P. and Scribner, J.D. (1983) *Journal of the American Chemical Society*, 105, 349-354.
39. Frecer, V. and Miertus, S. (1989) *Neoplasma*, 36, 257-272.
40. Friedman, L. (1970) in Carbonium Ions, Olah, G.A. and Schleyer, P.v.R., eds., Wiley-Interscience. New York, Vol.2, pp. 655-713.
41. Park, K.K., Archer, M.C., and Wishnok, J.S. (1980) *Chemico-Biological Interactions*, 29, 139-144.
42. Scribner, J.D. and Ford, G.P., (1982) *Cancer Letters*, 16, 51-56.
43. Ford, G.P. and Smith, C.T. (1987) *Journal of the Chemical Society, Chemical Communications*, 44-45.

44. Hammond, G.S. (1955) *Journal of the American Chemical Society*, 77, 334-338.
45. Singer, B. (1976) *Nature*, 264, 333-339.
46. Singer, B. (1979) *Journal of the National Cancer Institute*, 62, 3129-1339.
47. Miertus, S. and Trebaticka, M. (1984) *Journal of Theoretical Biology*, 108, 509-517.
48. Lawley, P.D. and Wallick, C.A. (1957) *Chem. Ind.*, 633.
49. Reiner, B. and Zamenhof, S. (1957) *Journal of Biological Chemistry*, 288, 475-486.
50. Lawley, P.D. (1961) *Journal de Chimie Physique*, 1011-1020.
51. Lawley, P.D. (1957) *Biochimica et Biophysica Acta*, 26, 450-451.
52. Loveless, A. and Hampton, C.L. (1969) *Mutation Research*, 7, 1-12.
53. Loveless, A. (1969) *Nature*, 223, 206-207.
54. Goth, R. and Rajewsky, M.F. (1974) *Proceedings of the National Academy of Sciences*, 71, 639-643.
55. Newbold, R.F., Warren, W., Medcalf, A.S.C., and Amos, J. (1980) *Nature*, 283, 596-599.
56. van Zeeland, A.A., Mohn, G.R., Neuhauser, A., and Ehling, U.H. (1985) *Environmental Health Perspectives*, 62, 163-169.
57. Rudiger, H.W., Schwartz, U. Serrand, E., Stief, M., Krause, T., Nowak, D., Doerger, G., and Lehnert, G. (1989) *Cancer Research*, 49, 5623-5626.

58. Isowa, G. Ishizaki, K., Sadamoto, T., Tanake, K., Yamaoka, Y., Ozzawa, K., and Ikenaga, M. (1991) *Carcinogenesis*, 12, 1313-1317.
59. Horsfall, M.J., Gordon, A.J.E., Burns, P.A., Zielenska, M., van der Vliet, G.M.E., and Glickman, B.W. (1990) *Environmental and Molecular Mutagenesis*, 15, (1990) 107-122.
60. Trey, J.E. and Gerson, S.L. (1989) *Cancer Research*, 49, 1899-1903.
61. Rossi, S.C., Conrad, M., Voight, J.M., and Topal, M.D. (1989) *Carcinogenesis*, 10, 373-377.
62. Citti, L., Mariana, L., Catapano, C.V., and Malvidi, G. (1989) *Annali dell'Istituto Superiore di Sanita*, 25, 5-9.
63. Schwartz, J.L., Turkula, T., Sagher, D., and Strauss, B. (1990) *Carcinogenesis*, 10, (1989) 681-685.
64. Guttenplan, J.B. (1990) *Mutation Research*, 233, 177-187.
65. Parthasarathy, R. and Fridley, S.M. (1986) *Carcinogenesis*, 7, 221-227.
66. Taylor, R. and Kennard, O. (1982) *Journal of Molecular Structure*, 78, 1-28.
67. Duncan, R.H. and Davies, G.S. (1989) *Journal of Theoretical Biology*, 140, 345-354.
68. Leonard, G.A., Thomson, J. Watson, W.P., and Brown, T. (1990) *Proceedings of the National Academy of Sciences*, 87, 9573-9576.
69. Ludlum, D.B. (1970) *Journal of Biological Chemistry*, 245, 477-482.

CHAPTER 2

Computational Methods

2.1 Transition State Optimization

The goal of this study is the computational characterization of the transition states for the reactions of the methyl-, ethyl-, and propyldiazonium ions, and their corresponding carbocations, with the O⁶ and, then, the N⁷ positions of the nucleic acid base guanine. The mathematical isolation of a transition state can be a difficult and tricky task. Schlegel has written several reviews which describe the process of optimizing both *equilibrium geometries* and transition state structures and which outline some of the more common algorithms [1-3]. What follows here is a brief mathematical description of what a transition state is and how it might be optimized as a prelude to an overview of the computational methods used in this study.

The Born-Oppenheimer approximation allows the construction of a multi-dimensional potential energy surface as a function of nuclear positions [4]. On this surface, there may be several minima which correspond to equilibrium structures. Given two such minima, it is possible to describe several paths between the minima. Each of these paths will pass through its own maximum. The transition state between the equilibrium structures associated with the two minima can be defined as the lowest of these maxima [1]. In other

words, a transition state is represented as a first-order saddle point on the potential surface. A first-order saddle point is a local maximum in one and only one direction and a local minimum in all perpendicular directions [1,2]. The direction along which the point is a local maximum corresponds to the reaction path.

If one wants a method capable of a meaningful transition state study, one looks for a method capable of, basically, three things. First, the method should be able to satisfactorily represent the potential energy surface. That is, it should be capable of accurate energy calculations for equilibrium geometries, transition state structures, and the surface connecting them for the chemical system under study. Methods will differ in their relative accuracies for various chemical systems and, especially for semiempirical methods parameterized to reproduce experimental parameters for stable species, the calculation of the transition state may be more difficult than the calculation of equilibrium structures. This means that, second, the method should have a suitable algorithm which allows optimization to the transition state from some starting conformation. Various algorithms are available [1]. Third, the method should provide a way of confirming that an optimized geometry is, indeed, a transition state. This may be accomplished by diagonalization of the multi-dimensional matrix of second derivatives of the energy with respect to position, the Hessian, to obtain the set of force constants. A transition state will have one, and only one, negative force constant.

In general, the process of optimizing transition states is more difficult than minimization. While minimization is usually ensured

if the starting coordinates are anywhere within the potential energy well of the minimum, transition state optimization requires a starting geometry which lies between the desired transition state and the multidimensional surface of inflection [5].

2.2 Semiempirical Calculations with MOPAC

The semiempirical method used in this study is MOPAC [6], a program which incorporates many individuals' work on various methods and algorithms into a general-purpose, easy-to-use molecular orbital package [5]. The appeal of the semiempirical approach is due primarily to its simplicity and speed without sacrificing accuracy. Rigorous traditional *ab initio* calculations on guanine-sized systems would require much more computer time than semiempirical methods. Using the larger basis sets necessary to reasonably ensure accurate results would easily render such an *ab initio* approach computationally prohibitive. Semiempirical methods have been shown to yield results comparable to those of *ab initio* calculations for systems which are included in or related to the set of compounds for which the semiempirical method has been parameterized [7,8]. Stewart has written an excellent article which outlines how MOPAC works [5] as well as a set of two articles describing the newest MNDO-PM3 (modified neglect of diatomic overlap-parametric method 3) parameterization and how it compares to the older MNDO (modified neglect of diatomic overlap) and AM1 (Austin model 1) options [7,9]. The following four paragraphs give an overview of MOPAC as it relates to this study with information taken from these articles by Stewart as well as from the MOPAC

manual. (The older MINDO/3, modified intermediate neglect of differential overlap, version 3, was not used in this study and will not be discussed here.) [6].

Each of the three semiempirical Hamiltonians within MOPAC used in this study makes a number of approximations. First, each atom is represented by a restricted basis set of one "s" orbital and three "p" orbitals. Second, all overlap integrals in the secular equation involving the overlap of two different atomic orbitals are neglected. Thus, the overlap matrix $\{S_{ij}\}$ becomes a unit matrix. Additionally, all two-electron integrals involving overlap of two atomic orbitals on different centers are ignored. Thus, no three- or four-center integrals are included.

The values of the remaining integrals in a MOPAC molecular orbital calculation are obtained through the parameterization either as parameters themselves (e.g., the PM3 one-center, two-electron integrals) or through other parameters which are part of various functional forms (e.g., the atomic orbital exponents). The parameters are optimized for a given set of compounds so as to best reproduce four gas-phase molecular properties -- heats of formation, dipole moments, ionization potentials, and molecular geometries. The basis for comparison is, in most cases, experimental values for these four properties, but, occasionally, high-level *ab initio* calculation are used. The result of the parameterization is a numerical value of each parameter for each included element. In MOPAC 5.0, MNDO is parameterized for 20 elements (H, Li, B, C, N, O, F, Al, Si, P, S, Cl, Cr, Zn, Ge, Br, Sn, I, Hg, and Pb) AM1 for 10 elements (H, B, C, N, O, F, S, Cl, Br, and I) and PM3

for 12 elements (H, C, N, O, F, Al, Si, P, S, Cl, Br, and I).

MOPAC uses its restricted basis set and optimized parameters in an SCF calculation which yields a density matrix, P , and a Fock matrix, F . These are used, along with the two-center, one-electron matrix, H , to obtain the total electronic energy as

$$E_{\text{elect}} = (1/2) \sum_i \sum_j P_{ij} (H_{ij} + F_{ij})$$

where i and j are atomic orbital indices. The heat of formation is then given by

$$\Delta H_f = E_{\text{elect}} + E_{\text{nuc}} + \sum_A E_{\text{el}}(A) + \sum_A \Delta H_f(A)$$

with $E_{\text{nuc}} = \sum \sum_{A < B} E_N(A, B)$ = core-core repulsion term for nuclei A and B

$E_{\text{el}}(A)$ = energy required to ionize the valence electrons of atom A

$\Delta H_f(A)$ = heat of atomization of atom A

$E_{\text{el}}(A)$ is calculated semiempirically while $\Delta H_f(A)$ is taken from experimental work.

The accuracy of this sort of semiempirical calculation depends on how good the optimized parameters are and on the program's theoretical framework which uses the parameters. The theoretical frameworks of MNDO, AM1, and PM3 are similar except in the core-core repulsion calculation where AM1 and PM3 introduce an additional term to reduce core-core repulsions beyond bonding distances. The big difference is in the parameterization. The PM3 parameterization is more extensive and allows more flexibility in reproducing experimental properties. Several parameters which are assigned from experimental results in the MNDO and AM1 methods

are, instead, included in the optimization in PM3. Also, while the slower parameterizations of MNDO and AM1 restricted the practical size of the parameterization sets to a few tens of compounds for MNDO and slightly more than one hundred for AM1, a more recently-developed method of parameterization allowed several hundred compounds to be considered in the PM3 parameterization. The idea is to allow the PM3 method to apply more reasonably to a wider range of molecules. References 5 and 9 tabulate MNDO, AM1, and PM3 results of various molecular properties for many different compounds along with experimental values for comparison. Each method has its strengths and weaknesses. However, in general, AM1 errors in heats of formation are about 40% less than those of MNDO, and PM3 errors are, again, about 40% less than those of AM1. Additionally, PM3 results for compounds containing hypervalent atoms are considerably better [5].

2.3 MOPAC 5.0 Optimization Methodology

In general, a transition state optimization begins by finding an approximate geometry to use as a starting point. As mentioned before, the requirements for the closeness of the starting geometry to the actual transition state is more demanding than for a geometry minimization. The actual transition state is then found through refinement of the starting geometry within the framework of certain optimization criteria. Calculation of force constants then allows confirmation of the final geometry as a first-order saddle point.

This study uses two versions of MOPAC -- MOPAC 5.0 and

MOPAC 6.0 [10]. The newer version includes an eigenvector following routine using Baker's algorithm [11] which can be used for optimization of both transition states and equilibrium structures [10]. The starting geometries for the eigenvector following calculations reported in this study were taken from the structures optimized with MOPAC 5.0. Therefore, this section will describe the MOPAC 5.0 optimization methodology and the necessary extensions to eigenvector following will be explained in the next section. Some examples of typical MOPAC input files as well as some associated Gaussian 90 input files are shown in the Appendix.

This study used the default geometry optimizer in MOPAC 5.0, namely the Broyden-Fletcher-Goldfarb-Shanno method [12-15], to minimize all equilibrium structures. This method uses a conjugate gradient algorithm [1] the accuracy of which is evidenced by the favorable comparison between MOPAC calculations and *ab initio* and experimental results mentioned earlier [7,8].

Precision in MOPAC calculations is based mainly on two factors, the SCF criterion and the geometry optimization criteria [5]. The SCF criterion stops the SCF iterations when both of the following tests are satisfied: (1) the electronic energy (in eV) for one iteration differs from that of the previous iteration by less than the value of "SELCON" (an adjustable parameter) with the difference between any three consecutive iterations less than ten times SELCON, and (2) the density matrix for one iteration differs from that of the previous iteration by a preset limit which is a multiple of SELCON. The default value of SELCON is 0.00001 kcal/mole. There are several criteria for stopping geometry optimization.

Principally, the following calculated quantities must be sufficiently small: (1) the predicted change in geometry (default criteria = 0.0001 Å), (2) the projected decrease in energy (default criteria = 0.001 kcal/mole), (3) the gradient norm (= scalar of the gradient vector) (default criteria = 1.0 kcal/mole/Å), and (4) the difference in heats of formation on two successive cycles (default criteria = 0.002 kcal/mole) [6]. Precision can be increased by the user through various keywords in the data file. In this study, the keyword "PRECISE" was used in all reported calculations. This makes the optimization criteria more demanding, usually by a factor of 100 [6].

Starting geometries for transition state optimizations in this study were determined by use of a method developed by Dewar, Healy, and Stewart [16] (invoked by the keyword "SADDLE") applied to approximate maxima in the reaction coordinate profile apparent in the results of coordinate-driving type calculations. The coordinate-driving technique is based on the assumption that the reaction is largely dominated by a change in one coordinate [1]. This coordinate, the distance between a certain atom in reactant 1, atom A, and a certain atom in reactant 2, atom B, is incrementally decreased from some relatively large separation (six to ten angstroms) while all other coordinates are fully optimized at each fixed A-B distance. A plot of heats of formation versus corresponding A-B distances then gives a rough reaction coordinate profile. In this study, the apparent maximum was selected by inspection and geometries chosen at point on either side of the maximum to use in a subsequent SADDLE calculation. The SADDLE calculation requires specification of the geometries of both the

reactants and the products of the reaction. Given a reaction $A \rightarrow B$, the "difference" between A and B is defined by Dewar, Healy, and Stewart [16] as

$$R = [\sum (a_i - b_i)^2]^{1/2}$$

where $A = \sum a_i$ define the geometry of A (a_i are the 3N-6 coordinates) and similarly for B. In the SADDLE calculation, R is decreased by some arbitrary amount (usually 5%) and the geometry of lower energy (either reactants or products) is optimized subject to the constraint $R_{\text{new}} = 0.95R_{\text{old}}$. This is repeated again and again with the system of lower energy being re-optimized each time since it is normally further from the transition state. R is iteratively reduced until, ideally, the reactants and products become identical at the transition state. The method fails, however, very close to the saddle point and the calculation is therefore stopped at some preset small R and further refinement with a gradient method is recommended [16]. In this study, refinement was accomplished by use of a non-derivative, non-linear least squares method due to Bartels which minimizes the gradient norm to arrive at the transition state. The keyword "NLLSQ" activates within MOPAC Bartels's algorithm. [6,17].

MOPAC then allows confirmation of the transition state through force constant calculation by use of the keyword "FORCE". A transition state, again, will have exactly one negative force constant. Along with force constant values, a description of the corresponding vibrations is provided in the MOPAC output which allows the user to determine which atoms contribute most significantly to each mode. A recent study of MNDO, AM1, and PM3

vibrational frequencies as compared with experimental values [18] reported that, for most systems, AM1 and PM3 results are superior to those of MNDO. The main failings of PM3 are for S-H and P-H stretching frequencies and, to a lesser extent, O-H stretches. AM1 fails most with ring and heavy-atom stretching frequencies. These factors will be considered again later to help in determining whether to use MNDO, AM1, or PM3 for the guanine alkylation transition states.

2.4 Eigenvector Following in MOPAC 6.0

MOPAC 6.0 includes an eigenvector following option which provides an alternative to both the BFGS minimization algorithm and Bartels's non-linear least squares procedure in that eigenvector following may be used to optimize either equilibrium structures or transition states [10]. The MOPAC manual describes the eigenvector following routine as appearing to be much faster than BFGS for minimization and also much faster and more reliable than Bartels's method for transition state optimization. The eigenvector following routine in MOPAC 6.0 uses a quasi-Newton Raphson algorithm due to Baker [11]. The goal of the algorithm is, of course, to locate a stationary point on the molecular potential energy surface which has the desired local surface characteristics -- all positive Hessian eigenvalues for a minimum or exactly one negative eigenvalue for a transition state. Baker's algorithm is based on earlier work by Banerjee *et. al.* [19].

Baker's analysis begins with the Taylor series expansion of the energy, E , about a point x_0 on the multidimensional energy surface

as

$$E(\mathbf{x}_0 + \mathbf{x}) = E(\mathbf{x}_0) + \mathbf{g}^+ \mathbf{x} + (1/2) \mathbf{x}^+ \mathbf{H} \mathbf{x} + \dots$$

where \mathbf{g} is the gradient vector at \mathbf{x}_0 , \mathbf{H} is the Hessian matrix at \mathbf{x}_0 , \mathbf{x} is a vector which gives the displacement from \mathbf{x}_0 , and the superscript $+$ indicates transposition. The Newton-Raphson method allows truncation of the series after the quadratic term which, after applying the stationary point condition $dE/d\mathbf{x} = 0$, leads to [19]

$$\mathbf{H} \mathbf{x} + \mathbf{g} = 0$$

Again following the analysis of Banerjee *et. al.*, the displacement variables are transformed via a unitary matrix which diagonalizes the Hessian matrix as $\mathbf{U}^+ \mathbf{H} \mathbf{U} = \mathbf{h}$ where \mathbf{h} is the vector of Hessian eigenvalues. Then, a new \mathbf{x} and a new \mathbf{g} are assigned as $\mathbf{x}_{\text{new}} = \mathbf{U}^+ \mathbf{x}_{\text{old}}$ and $\mathbf{g}_{\text{new}} = \mathbf{U}^+ \mathbf{g}_{\text{old}}$ so that the steps x_i and gradients g_i are associated with the eigenmodes. Then, the Newton-Raphson step is given by [19]

$$x_i = -g_i/h_i$$

As Banerjee *et. al.* pointed out, this last equation shows that the Newton-Raphson step is opposite the gradient for modes associated with positive Hessian eigenvalues and along the gradient for modes associated with negative Hessian eigenvalues. In other words, it tries to minimize along modes with positive Hessian eigenvalues and maximize along modes with negative eigenvalues. This is ideal for transition state searches if the starting geometry is in a region of the energy surface where the Hessian has one negative eigenvalue. The Newton-Raphson step would then maximize along that mode and minimize along all others to arrive at the

transition state. However, if the starting geometry is not close enough to the transition state geometry, then the Newton-Raphson step is inappropriate [11].

Baker [11] cites a method due to Poppinger [20] which, in the event of a transition state search begun in a region of the energy surface which does not have the necessary local structure, gives a way to correct towards the transition state. Poppinger suggested that if the Hessian had all positive eigenvalues, the mode associated with the lowest eigenvalue should be followed uphill, while if the Hessian had more than one negative eigenvalue, the mode associated with the least negative eigenvalue should be followed downhill. Baker pointed out the drawback in efficiency of this method due to acting along only one mode at a time.

Baker developed an algorithm which allows transition state optimization even if the starting geometry is in the wrong region of the energy surface. The algorithm is based on earlier work by Cerjan and Miller [21] as developed by Banerjee *et. al.* [19]. The method begins by calculating the gradient vector at the starting point. Then, an initial Hessian is assigned. The default Hessians in MOPAC 6.0 are a diagonal matrix for a minimum search or a numerically-determined Hessian for a transition state optimization. Other options are also available [10]. The next step is the diagonalize the Hessian (to determine the local surface structure) and to transform the gradient vector into the associated eigenmodes as described above for the Newton-Raphson procedure.

Now, instead of simply taking the Newton-Raphson step, Baker's algorithm takes a step of the form

$$x_i = -g_i/(h_i - \lambda)$$

where λ is a shift parameter on the Hessian eigenvalues [11]. Baker determines λ after the method of Banerjee *et. al.* [19]. For transition state searches, two shift parameters are used -- one for modes being minimized with respect to the energy and the other for modes being maximized. Within MOPAC 6.0, the maximum step thus found is limited to some user-supplied value (default = 0.20 angstroms or radians). Successive steps are made until suitable convergence (default is when largest gradient is less than 0.4 or 0.01 with "PRECISE").

Baker reported results of Gaussian 82 calculations on selected chemical systems using this method. He found that the method converged in fewer cycles than the standard transition state routine in Gaussian 82. In one case, with a starting geometry close to an energy minimum, Baker's algorithm successfully isolated a transition state structure while Gaussian 82's standard routine failed to converge. Baker also found the method suitable for energy minimization [11].

2.5 Density-Functional Theory

Density-functional theory (DFT) had its beginnings in the 1920s with the work of Thomas and Fermi [22-25] who developed an approximate atomic electron distribution from statistical considerations [26]. This led to the energy of the atom in terms of the associated electron density. However, since the method was not very accurate for atomic energies and since it failed for molecules, in that no molecular binding was predicted [27], the work of Thomas

and Fermi was not pursued much further until the 1960s when the work of Hohenberg, Kohn, and Sham began the development of modern density-functional theory. An excellent account of the development of DFT and its usefulness for understanding atoms and molecules is available in the book by Parr and Yang [26]. Much of the following description is adapted from their book.

In time-independent wave-function theory, the object is to solve the Schrodinger equation $H\Psi = E\Psi$ to obtain the electronic energy, E , and the many-electron wave function, Ψ with

$$H = \sum_i (-1/2) \nabla_i^2 + \sum_i v(r_i) + \sum_{i < j} (1/r_{ij})$$

where the summations are over all electrons and $v(r_i)$ is the potential acting on electron i due to the nuclei as

$$v(r_i) = -\sum_{\alpha} (Z_{\alpha}/r_{i\alpha})$$

where Z_{α} is the charge on nucleus α and the summation is over all nuclei [26].

The interelectron repulsion term involving $1/r_{ij}$ prevents an exact analytical solution for systems with more than one electron. Thus, some approximation must be used. The Hartree-Fock method approximates the total electronic wave function, Ψ , as an antisymmetrized product, a Slater determinant, of spin orbitals each of which is a combination of a spatial part and a spin part. In practice, the spatial part is expressed as a linear combination of one-electron basis functions [30]. The variation method is then applied to find the best Ψ by minimizing the energy with respect to the coefficients in the basis set expansion. The larger this basis set

of functions the more flexibility is available in representing Ψ and, thus, the more correct is the energy calculated from Ψ . However, even with an infinitely flexible basis set (an unattainable option), an exact determination of Ψ and E is not possible at the Hartree-Fock level due to electron correlation. Electron correlation effects are accounted for by using multiple determinants to represent Ψ , a technique called configuration interaction (CI), or by using a perturbation method, most commonly the Møller-Plesset (MP) method. Theoretically, with an infinitely flexible basis set and full CI, an exact Ψ and E could be determined [31]. However, this is not practical. The inclusion of electron correlation is computationally costly. Therefore, approximations are commonly accepted by limiting the extent of the CI or the level of the MP correction.

In DFT, the electron density $\rho(r)$, instead of the wave function, is the basic variable. In 1964, Hohenberg and Kohn [28] showed that electron density completely and uniquely determines the ground state of a system [26]. Hohenberg and Kohn developed an energy variational principle involving electron density which is analogous to the variational principle in wave function theory. Given a non-zero trial electron density which integrates to the total number of electrons, N , the energy calculated from the trial density is an upper bound to the of the true ground state energy [26]. The variational principle requires the ground state electron density to satisfy the stationary requirement

$$\delta\{E[\rho(r)] - \mu[\int \rho(r)dr - N]\} = 0$$

where $E[\rho(r)]$ denotes the energy as a functional of the electron density, and μ , the chemical potential, arises as a Lagrange

multiplier associated with the necessary constraint that the electron density integrates to N as $\int \rho(r) dr = N$ [26]. The ground state electronic energy being minimized is

$$E[\rho(r)] = \int \rho(r) v(r) dr + F[\rho(r)]$$

where $F[\rho(r)] = T[\rho(r)] + V_{ee}[\rho(r)]$. The integral in this expression for $E[\rho]$ accounts for the interaction of nuclei and electrons, while $T[\rho]$ is the electronic kinetic energy and $V_{ee}[\rho]$ is the interelectron repulsion which can be written as $V_{ee}[\rho] = J[\rho] +$ a nonclassical term where $J[\rho]$ is the coulombic repulsion. The chemical potential is then $\mu = v(r) + \{\delta F[\rho(r)]/\delta \rho(r)\}$

Kohn and Sham [29] introduced the use of orbitals into DFT as a indirect way to get around the problems of determining explicit expressions for $T[\rho]$ and $V_{ee}[\rho]$. Instead of working with the exact kinetic energy and electron density as

$$T = \sum_i n_i \langle \psi_i | (-1/2) \nabla_i^2 | \psi_i \rangle$$

$$\rho(r) = \sum_i n_i \sum_s |\psi_i(r,s)|^2$$

(where ψ_i is a spin orbital and $0 \leq n_i \leq 1$ is the number of electrons in ψ_i), Kohn and Sham used the special case where only N orbitals have $n_i = 1$ and all the other orbitals have $n_i = 0$ so that

$$T = \sum_{i \rightarrow N} \langle \psi_i | (-1/2) \nabla_i^2 | \psi_i \rangle$$

$$\rho(r) = \sum_{i \rightarrow N} \sum_s |\psi_i(r,s)|^2$$

These equations hold exactly for a system of N noninteracting electrons, and such a system is part of the approach of Kohn and

Sham [26]. In this reference system, there is no interelectron repulsion and, therefore, the system can be described by a single antisymmetrized determinant involving the N occupied ψ_i , and the ground state electron density is given exactly by this last expression for $\rho(r)$ [26].

Now, a new term is defined -- the exchange-correlation energy denoted $E_{xc}[\rho]$. The exchange-correlation energy accounts for the difference between $T_s[\rho]$ and the true kinetic energy $T[\rho]$ as well as the nonclassical term in $V_{ee}[\rho]$. Then, the energy functional can be written as [26]

$$E[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho] + \int \rho(r)v(r)dr$$

and the chemical potential is

$$\mu = v_{eff}(r) + \{\delta T_s[\rho(r)]/\delta \rho(r)\}$$

with the effective potential $v_{eff}(r)$ being [26]

$$v_{eff}(r) = v(r) + \int [\rho(r')/|r - r'|]dr' + \{\delta E_{xc}[\rho(r)]/\delta \rho(r)\}$$

For a given effective potential, $\rho(r)$ is found by solving the N one-electron equations

$$[-(1/2)\nabla^2 + v_{eff}(r)]\psi_i = \epsilon_i\psi_i$$

and then

$$\rho(r) = \sum_{i=1}^N \sum_s |\psi_i(r,s)|^2$$

Since $v_{eff}(r)$ depends on $\rho(r)$, a self-consistent approach is necessary. A $v_{eff}(r)$ is found from a guessed $\rho(r)$. This $v_{eff}(r)$ is then used to find the ψ_i and, thus, a new $\rho(r)$, and so forth [26].

These last three equations are the Kohn-Sham equations [29].

The Kohn-Sham approach differs from the Hartree-Fock method in that in Kohn-Sham theory, the exchange-correlation effects are fully incorporated (albeit approximately, in practice) whereas, in Hartree-Fock, the effects have to be added on, usually through configuration interaction or perturbation methods. If the exchange-correlation effects were known precisely, then Kohn-Sham theory would yield $\rho(r)$ and E exactly [26].

The simplest and most common way to approximate $E_{xc}[\rho]$ is the local density approximation (LDA) proposed by Kohn and Sham [26,29]. The LDA assumes that the exchange-correlation effects in an infinitesimal volume element about a point is the same as if the electron density were constant everywhere else [26,32]. Then,

$$E_{xc}^{LDA}[\rho] = \int \rho(r) \epsilon_{xc}(\rho) dr$$

where $\epsilon_{xc}(\rho)$ is the exchange-correlation energy per particle of a uniform electron gas with density $\rho(r)$ [26]. The LDA amounts to assuming homogeneity in what is, in reality, a nonuniform electron distribution.

This study uses the density-functional techniques of DGauss as implemented in the UniChem package of Cray Research, Inc. [32]. The DGauss program name stands for Density-Gaussian [33]. DGauss makes use of the computational efficiency of Gaussian-type orbitals in that the molecular orbitals, the electron density, and the exchange-correlation potential are each represented as a linear combination of Gaussians as [32]

$$\psi_i = \sum_p c_{ip} g_p$$

$$\rho(r) = \sum_{i \rightarrow N} |\psi_i|^2 \approx \sum_j p_j g_j$$

$$v_{xc}(r) = \{\delta E_{xc}[\rho(r)]/\delta \rho(r)\} = \sum_k \mu_k g_k$$

(The electron density approximation above is due to Sambe and Felton [34,35]). The g_p ($p = 1, 2, \dots, N$) are a set of contracted Gaussian basis functions similar to that in the Hartree-Fock method. The g_j and g_k are auxiliary basis sets of Gaussian-type functions. DGauss offers a choice of several sets of functions for the orbital and auxiliary basis sets as well as two methods for non-local corrections to $E_{xc}[\rho]$ [32]. The specific options chosen in this study will be described in Chapter 4.

A primary advantage of DFT over traditional *ab initio* calculations is computational efficiency. The computational time required for Hartree-Fock calculations scales as n^4 , in principle, where n is the number of basis functions. When correlation effects are included, this may become n^5 to n^6 . However, DFT scales, in principle, as n^3 [32,36]. This reduced scaling means DFT can be applied to larger systems. One goal of this study is the successful application of DFT to a guanine-sized system.

References

1. Schlegel, H.B. (1987) in Ab Initio Methods in Quantum Chemistry -I, Lawley, K.P., ed., Wiley & Sons, pp. 249-285.
2. Schlegel, H.B. (1982) *Journal of Computational Chemistry*, 3, 214-218.
3. Schlegel, H.B. (1989) in New Theoretical Concepts for Understanding Organic Reactions, Bertran, J. and Csizmadia, I.G., eds., Kluwer Academic Publishers, p. 33-53.
4. Levine, I.N. (1983) Quantum Chemistry, Third Edition, Allyn and Bacon, pp. 313-318.
5. Stewart, J.J.P. (1990) *Journal of Computer-Aided Molecular Design*, 4, 1-105.
6. Stewart, J.J.P. (1988) MOPAC Manual (Fifth Edition) A General Molecular Orbital Package, Air Force Systems Command, United States Air Force.
7. Stewart, J.J.P. (1989) *Journal of Computational Chemistry*, 10, 209-220.
8. Dewar, M.J.S. and Storch, D.M. (1985) *Journal of the American Chemical Society*, 107, 3898-3902.
9. Stewart, J.J.P. (1989) *Journal of Computational Chemistry*, 10, 221-264.
10. Stewart, J.J.P. (1990) MOPAC Manual (Sixth Edition) A General Molecular Orbital Package, Air Force Systems Command, United States Air Force.
11. Baker, J. (1986) *Journal of Computational Chemistry*, 7, 385-395.
12. Broyden, C.G. (1970) *Journal of the Institute for Mathematics and Applications*, 6, 222-231.

13. Fletcher, R. (1970) *Computer Journal*, 13, 317-322.
14. Goldfarb, D. (1970) *Mathematics of Computation*, 24, 23-26.
15. Shanno, D.F. (1970) *Mathematics of Computation*, 24, 647-656.
16. Dewar, M.J.S., Healy, E.F., and Stewart, J.J.P. (1984) *Journal of the Chemical Society, Faraday Transactions 2*, 80, 227-233.
17. Bartels, R.H. (1972) *University of Texas, Center for Numerical Analysis, Report CNA-44*, Austin, TX.
18. Coolidge, M.B., Marlin, J.E., and Stewart, J.J.P. (1991) *Journal of Computational Chemistry*, 12, 948-952.
19. Banerjee, A., Adams, N., Simons, J., and Shepard, R. (1985) *Journal of Physical Chemistry*, 89, 52-57.
20. Poppinger, D. (1975) *Chemical Physics Letters*, 35, 550.
21. Cerjan, C.J. and Miller, W.H. (1981) *Journal of Chemical Physics*, 75, 2800.
22. Thomas, L.H. (1927) *Proceedings of the Cambridge Philosophical Society*, 23, 542-548.
23. Fermi, E. (1927) *Rend. Accad., Lincei*, 6, 602-607.
24. Fermi, E. (1928) *Z. Phys.*, 48, 73-79 (english translation in March 1975).
25. Fermi, E. (1928) *Rend. Accad., Lincei*, 7, 342-346.
26. Parr, R.G. and Yang, W. (1989) Density-Functional Theory of Atoms and Molecules, Oxford University Press, New York.
27. Teller, E. (1962) *Reviews of Modern Physics*, 34, 627-631.
28. Hohenberg, P. and Kohn, W. (1964) *Physical Review*, 136, B864-B871.

29. Kohn, W. and Sham L.J. (1965) *Physical Review*, 140, A1133-A1138.
30. Levine, I.N. (1975) Molecular Spectroscopy, Wiley & Sons, New York.
31. Hehre, W.J., Radom, L., Schleyer, P.v.R., and Pople, J.A. (1986) Ab Initio Molecular Orbital Theory, Wiley & Sons, New York.
32. Unichem Chemistry Codes - SG-5151 1.0 (1991) Cray Research, Inc.
33. Andzelm, J. (1991) in Density Functional Methods in Chemistry, Labanowski, J.K. and Andzelm, J.W., eds., Springer-Verlag, New York.
34. Sambe, H. and Felton, R.H. (1974) *Journal of Chemical Physics*, 61, 3862.
35. Sambe, H. and Felton, R.H. (1975) *Journal of Chemical Physics*, 62, 1122.
36. Borman, S. (1990) *Chemical & Engineering News*, April 9, 22-30.

CHAPTER 3

Preliminary Semiempirical Calculations

3.1 Introduction

There are some indications of the relative merits of MNDO, AM1, and PM3. Stewart's announcement of PM3 was accompanied by the publication of an article comparing the three methods for a wide range of compounds [1]. These results indicate that, on average, PM3 performs the best for heats of formation, bonds length errors are generally reduced with PM3 while some bond angle errors are increased, PM3 dipole moment errors are intermediate between those of MNDO and AM1, and errors in ionization potentials are less for PM3 than for the other two methods. Thus, PM3 seems to be a significant improvement in many cases. (This view was not universally held when PM3 was published [2].) However, here we are less interested in how well the methods perform on average than in how well they perform for diazonium ions, carbocations, the nucleic acid base guanine, and the reactions between these species. The semiempirical results given in this chapter, calculated with MOPAC 5.0, are reported to help establish some basis for discussion of the three methods for these systems. All MOPAC calculations in this study were performed on the VAX computers at the National Institute of Environmental Health Sciences (NIEHS). *Ab initio* calculations were done on the Multiflow computer at NIEHS. Initial

geometries for all MOPAC calculations were set up using the DRAW program which allows on-screen geometry editing and subsequent output to a MOPAC-readable data file [3].

3.2 Alkyldiazonium Ions

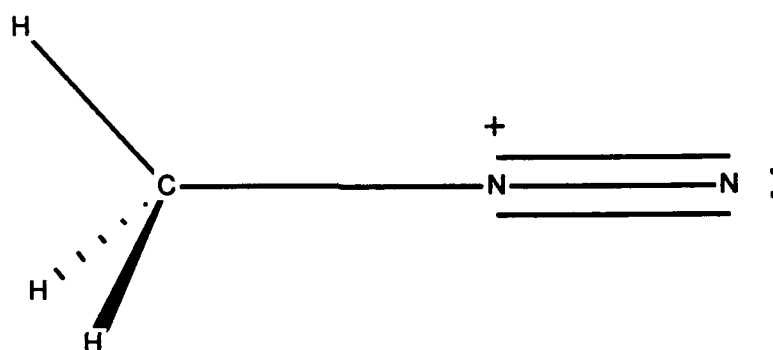
There is little experimental work on alkyldiazonium ions which have reported results easily comparable to calculations. Thus, the first comparisons in this study are made between MOPAC and *ab initio* calculations. The methyldiazonium ion (see Figure 3.1) was fully minimized using MNDO, AM1, and PM3. All bond lengths and bond angles were also optimized at the MP2/6-31G* level using Gaussian 90 [4]. The results are shown in Table 3.1. The C-H bond lengths were allowed to vary independently in the MOPAC calculations, and, therefore, the values shown are averages.

Table 3.1 - Geometry of the Methyldiazonium Ion

<u>Parameter</u>	<u>MNDO</u>	<u>AM1</u>	<u>PM3</u>	<u>MP2/6-31G*</u>
CH (Å)	1.112	1.131	1.105	1.091
CN (Å)	1.504	1.434	1.445	1.461
NN (Å)	1.107	1.109	1.105	1.128
HCN (°)	106.6	108.1	109.8	106.1
CNN (°)	180.0	180.0	180.0	180.0

When compared to the *ab initio* results, all three methods perform reasonably well. The most significant differences are in the C-N bond lengths, where the PM3 result is significantly superior, and the HCN bond angle, where the MNDO method prevails. Errors in bond angles may be less distressing than errors in bond lengths since bond stretching involves a larger force constant so that a

Figure 3.1 - Methyldiazonium Ion



small deviation in bond length is energetically more significant than a small deviation in bond angle. Also, the good performance of PM3 for the C-N length is encouraging since this is the bond being broken in an S_N2 -type alkylation by alkyldiazonium ions.

The breaking of this C-N bond is a significant event in both the S_N2 process, as mentioned above, and in the S_N1 mechanism as a means of generating carbocations. The dissociation of methyl- and ethyldiazonium ions was studied, using MNDO, by Ford and Scribner [5]. Some of their work is summarized in Table 3.2 along with similar results done as part of this study and some experimental values cited by Ford and Scribner.

When compared to the experimental values, the PM3 heat of formation is significantly better for the methyl cation but slightly inferior for the ethyl cation and the methyldiazonium ion. Although there is no experimental value for the ethyldiazonium ion, the results of the three methods do not differ drastically. For the enthalpy change for the methyldiazonium ion dissociation [found by $\Delta_{rxn}H = \Delta_fH(CH_3)^+ + \Delta_fH(N_2) - \Delta_fH(CH_3N_2^+)$ with MOPAC heats of formation used in the calculation], one experimental value is close to the AM1 result while the other experimental value is close to the PM3 result. Ford and Scribner attribute the relatively poor performance of MNDO to the error in the calculated heat of formation for the methyl cation. Interestingly enough, if the heat of formation of N_2 , which is overestimated by all three methods, is set arbitrarily to zero, then the AM1 and PM3 heats of reaction differ by only 0.3 kcal/mole (30.7 kcal/mole for AM1 and 30.4 kcal/mole for

**Table 3.2 - Diazonium Ion Enthalpic Calculations
(kcal/mole)**

<u>Parameter</u>	<u>MNDO</u>	<u>AM1</u>	<u>PM3</u>	<u>Expt.</u>
$\Delta_f H$ of CH_3^+	243.9	252.4	256.6	261.3 ²
$\Delta_f H$ of C_2H_5^+	219.7 219.7 ¹	216.8	222.5	216.0 ³
$\Delta_f H$ of CH_3N_2^+	223.7 223.5 ¹	221.7	226.2	223 ⁴ 209.4 ⁵
$\Delta_f H$ of $\text{C}_2\text{H}_5\text{N}_2^+$	214.0 213.8 ¹	211.5	217.2	-
$\Delta_{\text{rxn}} H$ for $\text{CH}_3\text{N}_2^+ \rightarrow$ $\text{CH}_3^+ + \text{N}_2$	28.5 28.4 ¹	41.9	48.0	38.3 ⁶ 51.9 ⁷
$\Delta_{\text{rxn}} H$ for $\text{C}_2\text{H}_5\text{N}_2^+ \rightarrow$ $\text{C}_2\text{H}_5^+ + \text{N}_2$	14.0 13.9 ¹	16.5	22.9	-

¹MNDO values from Reference 5

²Cited in Reference 5

³Cited in Reference 5 from Reference 6

⁴Cited in Reference 5 from Reference 7

⁵Cited in Reference 5 from Reference 8

⁶Based on CH_3N_2^+ heat of formation of 223 kcal/mole

⁷Based on CH_3N_2^+ heat of formation of 209.4 kcal/mole

PM3) while MNDO gives a result about 10 kcal/mole lower at 20.2 kcal/mole.

Although experimental values of these parameters are difficult to obtain and vary from one technique to another, these results seem to indicate that MNDO is the least satisfactory of the three methods for these systems.

A later calculation by Ford, however, gave Hartree-Fock (6-31G**//6-31G*) dissociation energies for dissociation into carbocations and diatomic nitrogen of 25.8 kcal/mole and 5.2 kcal/mole for the methyl- and ethyldiazonium ions respectively [9]. Comparisons of MOPAC and *ab initio* reaction energetics require some care. *Ab initio* calculations yield total energies, E , for vibrationless species at 0 K as $E = E_{\text{elec}} + E_{\text{nucI}}$ where E_{elec} is the electronic energy and E_{nucI} is the nuclear repulsion. *Ab initio* calculations can also provide a zero-point vibrational energy from normal mode vibrational frequencies via an harmonic oscillator approximation of the 0 K energy surface. The internal energy at any temperature above 0 K must include this zero-point energy as well as the rotational, vibrational, and translational contributions as calculated from the respective partition functions [10]. In MOPAC, the SCF "energy" is defined as the heat of formation at 25°C, and the semiempirical parameters are optimized to that end. This value implicitly includes the electronic energy, the zero-point energy, the rotational, vibrational, and translational contributions, as well as a pV term. Thus, a direct comparison of MOPAC's heat of formation to an *ab initio* energy is tenuous at best. However, it seems reasonable that in taking differences of MOPAC heat of formation

values, the zero-point energies as well as the rotational, vibrational, and translational energies would approximately cancel. Dissociation or association is analogous to a conformational change except that a vibrational mode has been lost to translation or *vice versa*. Table 7.45 in reference 11 shows zero-point energies of some small isomers differing by only a few kcal/mole. Enthalpy corrections for rotational, vibrational and translational motion at 300 K is typically about 2 kcal/mole for small molecules [Table 6.52 in reference 11]. The pV term may still contribute on the order of RT , but since $RT \approx 0.6$ kcal/mole at 25°C, this is not particularly significant. With this in mind, comparison of Ford's 6-31G**//6-31G* calculations [9] with the MOPAC results in Table 3.2 shows that these *ab initio* values agree more closely with MNDO. However, when the effects of electron correlation began to be included in the same study by Ford, the dissociation energies increased (see Table 3.3) and the comparisons with MNDO become less favorable.

Table 3.3 - *Ab Initio* Dissociation Energies of Alkyldiazonium Ions (Using the 6-31G Basis Set and HF/6-31G* Geometries)^{1,2}**
(kcal/mole)

<u>Type</u>	<u>HF</u>	<u>MP2</u>	<u>MP3</u>	<u>MP4(SDQ)</u>
methyl	25.8	46.0	42.3	42.4
ethyl	5.2	20.5	17.2	17.4

¹Adapted from Table IV in Reference 9

²Dissociation \Rightarrow $RN_2^+ \rightarrow R^+ + N_2$ (vibrationless species at 0 K)

3.3 Guanine

Table 3.4 gives MOPAC results for bond lengths in 9-methylguanine along with experimental values. Most of the bond lengths are slightly overestimated. The exceptions are the exocyclic bonds. However, all bond lengths are fairly well represented. These calculations provide little basis for deciding between MNDO, AM1, or PM3.

Table 3.4 - Bond Lengths in 9-Methylguanine (Å)

<u>Bond</u>	<u>MNDO</u>	<u>AM1</u>	<u>PM3</u>	<u>Expt.</u> ¹
N ¹ -C ²	1.396	1.410	1.414	1.375
C ² -N ³	1.335	1.355	1.339	1.327
C ² -N ²	1.415	1.414	1.423	1.341
N ³ -C ⁴	1.383	1.383	1.399	1.355
C ⁴ -C ⁵	1.413	1.442	1.407	1.377
C ⁵ -C ⁶	1.454	1.448	1.448	1.415
C ⁶ -O ⁶	1.221	1.239	1.216	1.239
C ⁶ -N ¹	1.450	1.422	1.453	1.393
C ⁵ -N ⁷	1.393	1.396	1.398	1.389
N ⁷ -C ⁸	1.336	1.345	1.341	1.304
C ⁸ -N ⁹	1.420	1.418	1.417	1.374
N ⁹ -C ^{1'}	1.463	1.424	1.461	1.476
Average Unsigned Errors	0.030	0.035	0.034	-

¹From Reference 12, p.52

Pedersen *et. al.* have reported *ab initio* (3-21G) calculations of the geometry and energy of an alkylation product, namely 9-methyl-O⁶-methylguanine [13]. There are two nearly energetically equivalent isomers -- the "distal" form with the O⁶

methyl group coplanar with the rings and oriented toward the hydrogen-bonding region, and the "proximal" form with the methyl group still coplanar with the rings, but at a dihedral angle differing by 180° (see Figure 3.2). The distal form was indicated in an x-ray determination of the structure of O⁶-methylguanosine [14]. However, since this would place the methyl group in a position to interfere with hydrogen-bonding in the double helix, Pedersen *et. al.* have suggested that proximal is the more likely biological form. Their results are compared in Table 3.5 with those of MOPAC and with x-ray values.

For both the distal and proximal forms, the only geometrical parameter for which PM3 gives better (as compared to *ab initio* and x-ray values) results than MNDO or AM1 is the dihedral angle δ . Even there, the difference is not particularly significant, especially considering the comparatively small force constants of dihedral angles. For each of the other geometrical parameters, the PM3 results are intermediate between those of MNDO and AM1 and comparable to each. AM1 appears to be superior for the R2 value which would be important to this study as the forming bond for O⁶ alkylation.

AM1 gives a heat of formation difference between the two forms which is closer to the corresponding *ab initio* energy difference value. However, comparing the actual heats of formation given by the three methods (Distal: MNDO, 7.93 kcal/mole; AM1, 62.32 kcal/mole; PM3, 14.95 kcal/mole; Proximal: MNDO, 10.64 kcal/mole; AM1, 65.70 kcal/mole; PM3, 15.03 kcal/mole) indicates a large difference between AM1 and the other two methods. Thus,

Figure 3.2 - Distal and Proximal Forms of
9-Methyl-O⁶-Methylguanine

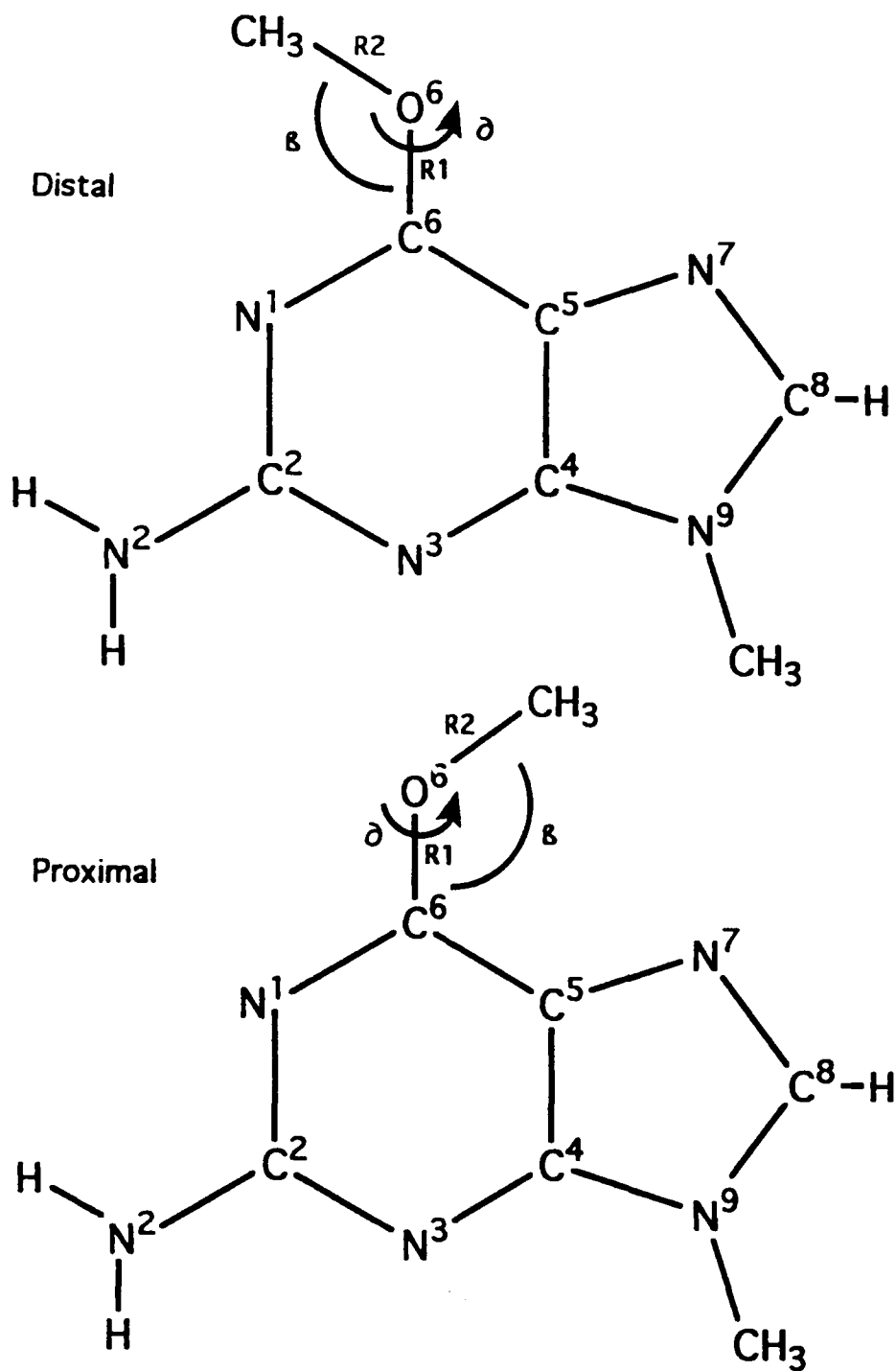


Table 3.5 - 9-Methyl-O⁶-Methylguanine

Distal					
Parameter	MNDO	AM1	PM3	3-21G ¹	X-ray ²
R1 (Å)	1.338	1.370	1.354	1.3323	1.338
R2 (Å)	1.410	1.430	1.414	1.4466	1.447
β (°)	124.4	117.8	118.3	119.47	116.4
δ (°)	1.7	1.2	0.9	0.0	0
E ³ (kcal/mole)	0.0	0.0	0.0	0.0	0

Proximal					
Parameter	MNDO	AM1	PM3	3-21G ¹	X-ray ²
R1 (Å)	1.341	1.372	1.350	1.3337	N/A
R2 (Å)	1.408	1.428	1.413	1.4430	N/A
β (°)	124.6	117.2	118.0	126.91	N/A
δ (°)	178.2	179.3	179.9	180.0	N/A
E ³ (kcal/mole)	2.71	3.38	0.08	4.61	N/A

¹From Reference 13

²From Reference 13

³To be consistent with Reference 13, the energies (enthalpies for MOPAC) of the distal form were set arbitrarily to zero and the values for the proximal form were reported relative to that zero.

AM1's apparent good performance for the energy difference may be fortuitous.

In Table 3.6, the hydrogen-bonding distances calculated for MNDO, AM1, and PM3 for a Watson-Crick G:C base pair are listed (see Figure 3.3 for parameter descriptions) along with results of *ab initio* calculations using the MINI-1 basis set (similar to STO-3G except that different exponents are used for "s" and "p" functions in

a given shell [15]) and x-ray determinations. In the MOPAC calculations, the geometry was fully optimized. In the MINI-1 calculation, the intramolecular geometries were held fixed at previously optimized values [16] and the intermolecular positions were optimized while keeping the two ring systems coplanar. The failure of MNDO to represent this hydrogen-bonding is evident. This has been corrected in AM1 and PM3 by the inclusion of an additional term in the core-core repulsion as mentioned earlier [17], and the AM1 and PM3 values are quite good with PM3 being, perhaps, slightly superior.

Table 3.6 - Hydrogen-Bonding Distances in G:C Pair

<u>Parameter</u>	<u>MNDO</u>	<u>AM1</u>	<u>PM3</u>	<u>MINI-1</u> ¹	<u>X-ray</u> ²
HB1 (Å)	3.97	3.06	2.81	2.96	2.91
HB2 (Å)	3.78	3.04	2.80	2.94	2.95
HB3 (Å)	3.95	3.08	2.85	2.91	2.86

¹From Reference 18

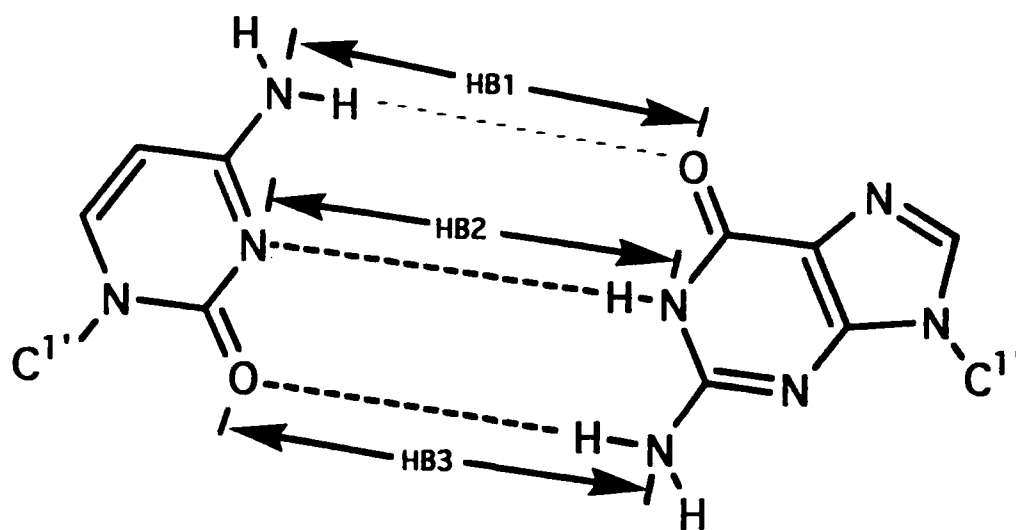
²From Reference 12

The binding enthalpy of a G:C base pair can be calculated as the difference between the heat of formation of the hydrogen-bonded pair and the heats of formation of the isolated bases.

$$\text{Binding Enthalpy} = \Delta_f H_{\text{G:C}} - \Delta_f H_{\text{G}} - \Delta_f H_{\text{C}}$$

The MOPAC results for this binding energy is 4.4 kcal/mole for MNDO, 13.4 kcal/mole for AM1, and 11.8 kcal/mole for PM3. An experimental binding energy value from temperature-dependent field ionization mass spectroscopy is 21.0 kcal/mole [19] which indicates, again, that AM1 and PM3 better represent hydrogen bonds.

Figure 3.3 - Hydrogen Bonding in G:C Base Pair



3.4 Methyldiazonium Ion - Formamide Transition State

Formamide was one of several model compounds used by Ford and Scribner in their MNDO study of alkyldiazonium ions [5].

Similarly, the current study also uses formamide to represent the O⁶ position of guanine in transition state calculations for the reaction between formamide and the methyldiazonium ion. This allows the application of fairly rigorous *ab initio* techniques to provide a reasonable basis for comparison with MOPAC results.

Before discussing the transition state calculations, it is useful to examine the results of calculations on the reactants and products. The methyldiazonium ion was discussed previously in the section on preliminary calculations. The results of full geometry optimizations of the other reactant, formamide, and the product of the reaction, a positively-charged O⁶-methylated formamide, are shown in Table 3.7 and 3.8. (Nomenclature for potentially ambiguously-named angles in Tables 3.7, 3.8, and 3.9 is indicated in Figure 3.4.)

When compared with the *ab initio* calculations, all three MOPAC methods yield reasonable results for formamide. In each case, a fairly planar molecule is predicted with MNDO showing the greatest out of plane deviation. The worst offense is MNDO's prediction of the C-N bond length which is overestimated by all three methods with AM1 giving the best value as compared to the MP2/6-31G* result. For the O⁶-methylated product, AM1 again does well except for the C-O bond length where the AM1 error is about twice that of MNDO. The PM3 geometry results are comparable to those of the other methods. PM3's larger errors are in bond angles.

Figure 3.4 - Methylation of Formamide by
Methyldiazonium Ion

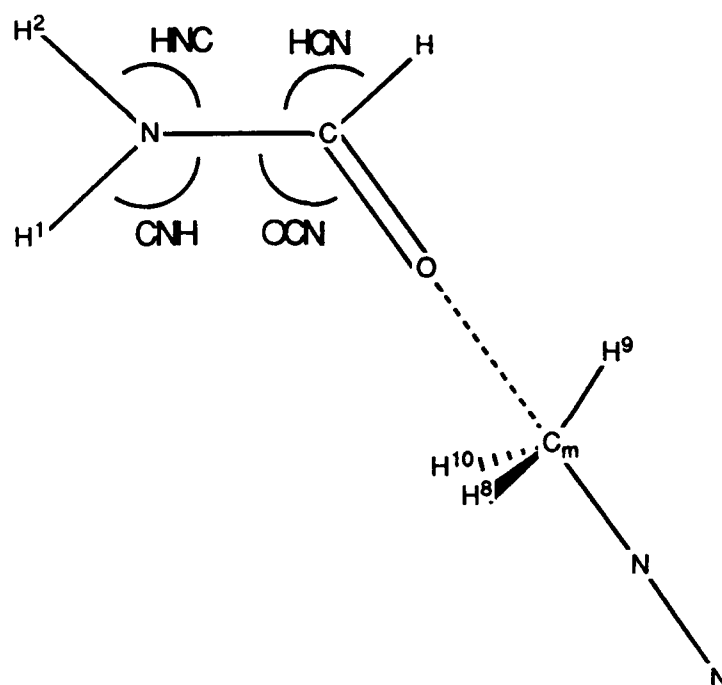


Table 3.7 - Results of Formamide Calculations

<u>Parameter</u>	<u>MNDO</u>	<u>AM1</u>	<u>PM3</u>	<u>MP2/6-31G*</u>
CO (Å)	1.225	1.243	1.220	1.224
CN (Å)	1.409	1.367	1.391	1.360
CH (Å)	1.108	1.114	1.101	1.104
NH1 (Å)	1.002	0.990	0.992	1.011
NH2 (Å)	1.000	0.986	0.990	1.008
OCN (°)	121.1	122.0	117.7	124.7
CNH (°)	117.8	120.6	121.6	118.9
HCN (°)	114.4	115.0	117.9	112.3
HNC (°)	115.4	121.2	120.5	121.9

Table 3.8 - Results of Calculations on
O⁶-Methylated Formamide (with +1 Charge)

<u>Parameter</u>	<u>MNDO</u>	<u>AM1</u>	<u>PM3</u>	<u>MP2/6-31G*</u>
CO (Å)	1.312	1.340	1.328	1.283
CN (Å)	1.337	1.318	1.324	1.303
CH (Å)	1.107	1.116	1.108	1.090
NH1 (Å)	1.008	1.005	0.994	1.019
NH2 (Å)	1.006	1.003	0.990	1.016
C _m O (Å)	1.433	1.451	1.428	1.474
OCN (°)	115.7	116.2	111.7	119.6
C _m OC (°)	125.1	116.5	118.0	119.6
CNH (°)	124.1	121.7	122.8	120.6
HCN (°)	120.7	122.5	124.9	119.2
HNC (°)	120.5	120.8	119.7	120.9
H8C _m O (°)	110.5	109.5	111.1	105.9
H9C _m O (°)	105.4	101.8	101.3	109.9
H10C _m O (°)	110.5	109.5	111.1	105.9

The procedure described in Chapter 2 for transition state optimization was applied in separate sets of calculations for MNDO,

AM1, and PM3. First, a coordinate-driving type calculation [20] was performed by incrementally closing the distance between the carbon of the attacking methyldiazonium ion and the oxygen of the formamide molecule with full optimization of all other geometrical parameters at each fixed carbon-oxygen distance. The initial attack was set up with the diazonium nitrogens, the diazonium carbon, and the formamide's carbon and oxygen atoms all being colinear and with the diazonium nitrogens trailing the carbon (see Figure 3.4). A plot of heat of formation versus C_m -O distance allowed approximate location of the energy maximum during the course of the reaction. In each case, the energy maximum was preceded by a minimum representing stabilization due to favorable interactions between the positively-charged ion and the partial negative charge of the formamide oxygen. Geometries were selected on either side of the maximum and used in a "SADDLE" calculation [21] to find the approximate transition state. The transition state was then refined using Bartels's method [22] and confirmed through the calculation of exactly one negative force constant. Various geometrical parameters of the transition state thus characterized are listed in Table 3.9 along with the results of *ab initio* optimizations (calculated with Gaussian 90's default transition state routine) at the HF/6-31G* level with analytic Hartree-Fock force constants computed at the initial step, and at the MP2/6-31G* level with numerically-determined force constants for active variables. The MOPAC calculations were full geometry optimizations. Each of the three semiempirical optimizations yielded a structure in which the heavy atoms were all essentially coplanar. The angles not shown in

Table 3.9 were not significantly changed from their values in formamide. In the *ab initio* calculations, only those parameters shown in the table were optimized while all others were held fixed at values chosen by averaging the results of the MNDO, AM1, and PM3 calculations. The HC_mO angle given is the average of the three HC_mO angles.

Table 3.9 - Transition State Calculations for the O-Methylation of Formamide by the Methyldiazonium Ion

<u>Parameter</u>	<u>MNDO</u>	<u>AM1</u>	<u>PM3</u>	<u>HF/6-31G*</u>	<u>MP2/6-31G*</u>
NH1 (Å)	0.999	0.997	0.993	0.996	1.013
CN (Å)	1.366	1.340	1.355	1.326	1.333
CO (Å)	1.247	1.276	1.254	1.217	1.249
C_mO (Å)	2.213	1.984	2.045	2.344	2.190
C_mN (Å)	2.078	1.843	1.871	1.757	1.738
NN (Å)	1.104	1.104	1.100	1.074	1.128
C_mOC (°)	155.5	124.2	129.7	157.2	132.1
HC_mO (°)	88.3	85.8	85.6	81.1	82.5

For the bonds which are "stable" (not being formed or broken) during the reaction, the three MOPAC methods are comparable to each other. Perhaps the most noteworthy of these is the C-N bond which is overestimated by all three methods with AM1 giving the best comparison with the *ab initio* and PM3 being intermediate between MNDO and AM1. However, the most significant comparisons to be made are for the bonds being formed or broken during the reaction, namely the $\text{C}_m\text{-O}$ and $\text{C}_m\text{-N}$ bonds. These represent bonds in transition, and, in addition to being important for transition state characterization, they are also the most likely bonds to stretch the

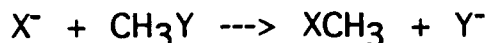
capabilities of semiempirical methods parameterized for stable species. The results for these bonds show that, when compared to the Hartree-Fock values, MNDO performs the best for the C_m -O bond but predicts the C_m -N length poorly. For AM1, the reverse is true. In both bonds, the PM3 result is intermediate between the other results. As some electron correlation effects are included (at the MP2 level) in the *ab initio* calculation, the comparison with PM3 becomes more favorable as MNDO and AM1 maintain significant deviations for the C_m -N and C_m -O bonds respectively while the PM3 deviations become relatively smaller. Also, the MP2/6-31G* results for the C_mOC angle compare favorably with PM3.

The calculated heats of formation for the methyldiazonium ion-formamide transition state are 179.7 kcal/mole for MNDO, 191.8 kcal/mole for AM1, and 192.1 kcal/mole for PM3. The agreement between AM1 and PM3 versus the lower value of MNDO may be due to the improved core-core repulsion treatment of AM1 and PM3 [17] which may aid in describing these bonds in transition which are longer than ordinary chemical bonds. Comparing these transition state heats of formation with those of the isolated reactants gives gas-phase heats of activation of 4.4 kcal/mole for MNDO, 2.6 kcal/mole for AM1, and 6.3 kcal/mole for PM3. From the MP2/6-31G* vibrationless energies, the *ab initio* activation energy was calculated to be -14.5 kcal/mole. This discrepancy may be genuine, or it may be due to the fact that all the MOPAC calculations were full geometry optimizations while, in the *ab initio* transition state calculation, the active variables were limited to those

parameters in Table 3.9. In any case, since the three MOPAC methods gave activation enthalpies in reasonable agreement with each other, it was decided that a full *ab initio* geometry optimization would not be accomplished since it would be of limited benefit in deciding which MOPAC method was the best for this study. Instead, in the next section, energy trend information provided by PM3 calculations is compared with *ab initio* work from the literature to further investigate the performance of MOPAC. A more complete comparison of the energetics predicted by the various computational methods for this and the analogous guanine reaction will be given in Chapter 5 when the DGauss results are included.

3.5 Typical S_N2 Transition States

In 1981, Wolfe and Mitchell reported the results of their *ab initio* (4-31G) calculations in the study of several simple S_N2 reactions [23,24]. Their work gave some geometry and energy relationships which can easily be used as a basis for comparison with MOPAC calculations. The reactions they studied were of the form



with $Y = F$ or OH and, for each Y , $X = H, H_2N, HO, HCC, CH_3O, F, NC, HOO, CN,$ and FO . They calculated the geometries and energies of several points along the reaction path, namely, the isolated reactants (I), the pre-transition state ion-molecule complex (II), the transition state (III), the post-transition state ion-molecule complex (IV), and the isolated products (V).

Wolfe and Mitchell plotted the transition state values of the C-Y distance and, then, the YCH angle against the ΔE for the reaction using the 4-31G energies [$\Delta E = E(V) - E(I)$] for the full set of X's for each Y. The resulting four plots were each fairly linear with positive slopes for C-Y length versus ΔE plots and negative slopes for the YCH angle versus ΔE plots. In other words as the reactions became more exothermic, the transition state C-Y distance decreased while the YCH angle increased. This is in keeping with what Wolfe and Mitchell call the Bell-Evans-Polanyi-Leffler-Hammond effect which says that the more exothermic a reaction is, the more the transition state is like the reactants [23,25]. Linear regression analysis gave r values of 0.952 for the C-F distance plot, 0.976 for the FCH angle plot, 0.967 for the C-O distance plot, and 0.994 for the OCH angle plot.

In this study, PM3-optimized results were used to construct similar plots except that the enthalpy change of the reaction was used in place of ΔE . The plots are shown in Figures 3.5 and 3.6. The overall trends are fairly linear with the plot for the FCH angle being the worst linear fit. (One less point is plotted for the reactions with CH_3OH because the reaction with H^+ led to an undesirable -C-H-H interaction.)

Energetically, the PM3 results are also qualitatively similar to the 4-31G calculations of Wolfe and Mitchell. Each transition state was preceded by an ion-molecule complex which was lower in energy than either the transition state or the isolated reactants. This ion-molecule energy well is consistent with a reaction rate which is less than the collision rate even when the transition state is lower

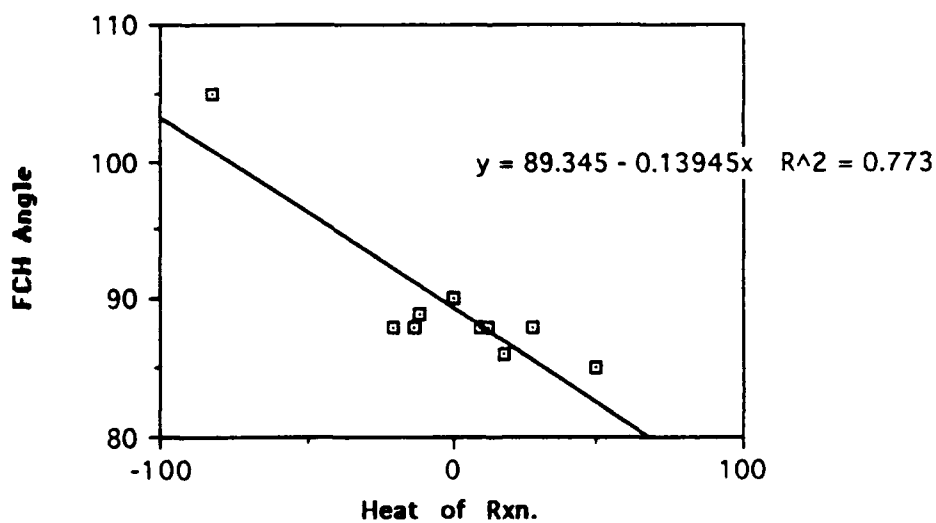
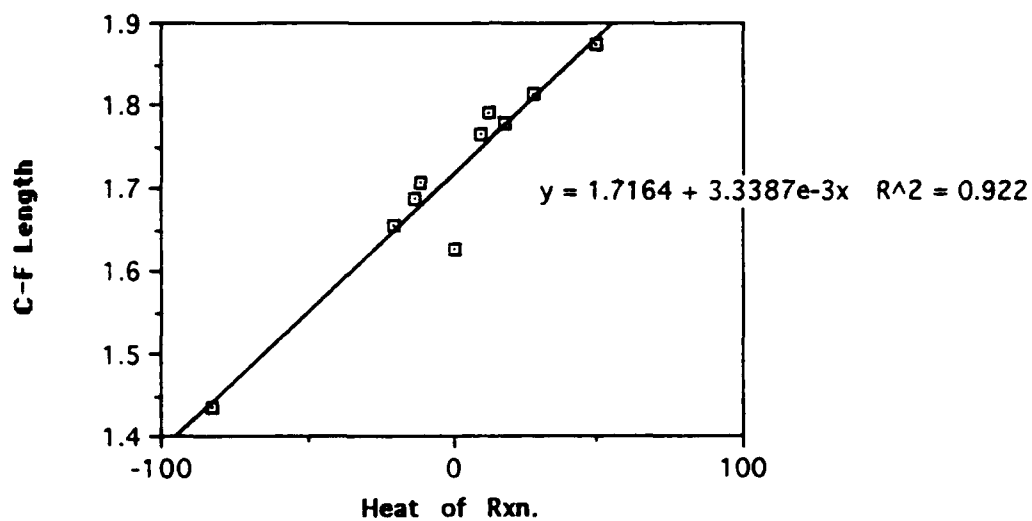
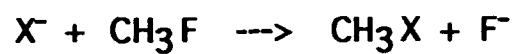
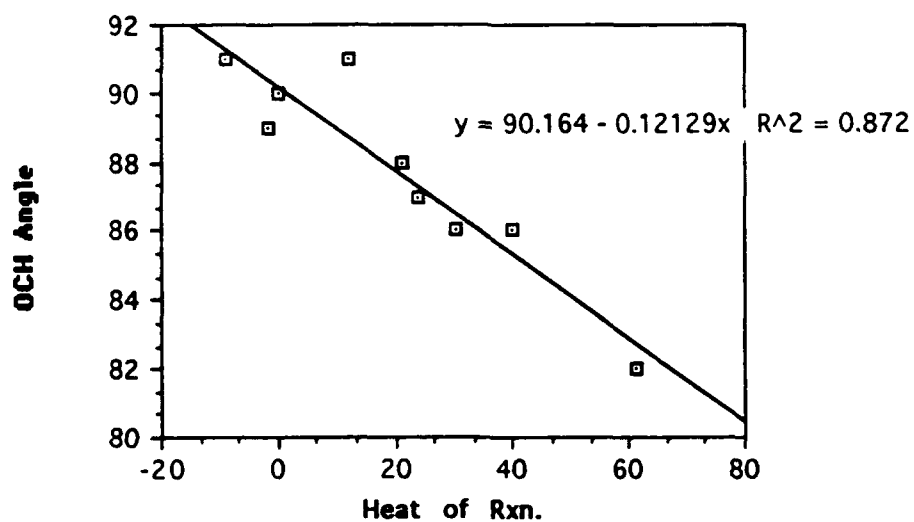
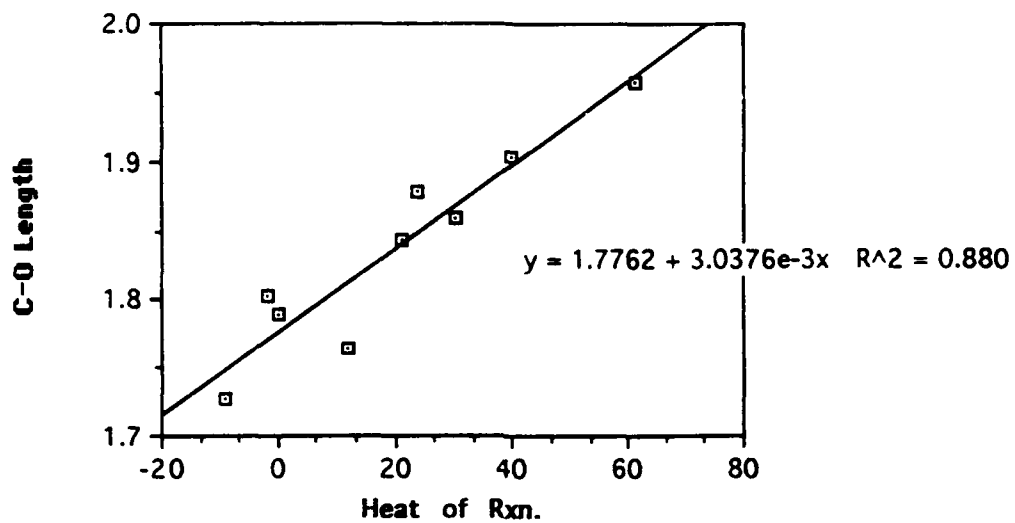
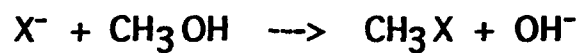
Figure 3.5 - Plots of Geometry versus $\Delta_{\text{rxn}}H$ for

Figure 3.6 - Plots of Geometry *versus* $\Delta_{\text{rxn}}H$ for

in energy than the isolated reactants [24,26]. The energy difference between the ion-molecule complex and the transition state, the "intrinsic barrier", has been shown to correlate with reaction efficiency [24,27]. Wolfe and Mitchell reported 4-31G intrinsic barriers for several degenerate S_N2 reactions where the attacking group and the leaving group are the same (i.e., $X = Y$) [24]. Their results are listed in Table 3.10 along with intrinsic barriers calculated with PM3. For polyatomic ions, italics indicate which atom is attacking the methyl carbon. The comparisons are not very good, quantitatively, in several cases. However, with the exception of $X = \text{HOO}$ and $X = \text{F}$, the ordering of the barriers for PM3 is the same as the 4-31G ordering.

Table 3.10 - Intrinsic Barriers for Degenerate S_N2 Reactions (kcal/mole)

<u>X(=Y)</u>	<u>4-31G¹</u>	<u>PM3</u>
HCC	50.4	57.4
NC	43.8	47.9
CH ₃ O	23.5	40.0
HO	21.2	31.3
HOO	18.5	47.7
HS	15.6	23.9
F	11.7	45.1
Cl	5.5	9.7

¹From Reference 24.

3.6 Conclusions from Preliminary Calculations

The preliminary calculations in this chapter do not clearly show which MOPAC Hamiltonian, MNDO, AM1, or PM3, can be expected to give the most reliable results for the transition state calculations on guanine. MNDO compares favorably for some parameters but did not perform as well for the C-N bond length in the methyldiazonium ion, the heat of formation of the methyl cation, the heat of reaction for methyldiazonium ion dissociation, or the hydrogen-bonded base pair interactions. AM1 and PM3 results are more comparable. However, PM3 seems to be the best "jack-of-all-trades" method. It does not always give the best comparison with experimental work or *ab initio* calculations, but, when it "loses", the margin is usually small. This was shown also in the formamide transition state calculations. The other two methods, MNDO and AM1, show larger errors more often.

Another consideration is that since the size of the PM3 parameterization set is several times that of AM1 or MNDO [28], it may be expected that PM3's broader database may make the extension to transition states easier. In addition, the relative failure of AM1 to properly calculate ring stretching frequencies (as mentioned earlier [29]) may indicate that PM3 will better represent the potential surface for a system involving guanine. Finally, the comparisons of PM3 results with those of *ab initio* work on typical S_N2 reaction energetics showed that PM3 can be expected to give useful energy trend information. Therefore, PM3 will be used for the calculations involving guanine.

References

1. Stewart, J.J.P. (1989) *Journal of Computational Chemistry*, 10, 221-264.
2. Dewar, M.J.S., Healy, E.F., Holder, A.J., and Yuan, Y. (1990) *Journal of Computational Chemistry*, 11, 541-542.
3. Storch, D.M. (1989) DRAW: Molecular Drawing Program, Version 2.00, Quantum Chemistry Program Exchange, Department of Chemistry, Indiana University, Indiana, 47405, No. 493.
4. *Gaussian 90*, M.J. Frisch, M. Head-Gordon, G.W. Trucks, J.B. Foresman, H.B. Schlegel, K. Raghavachari, M.A. Robb, J.S. Binkley, C. Gonzalez, D.J. Defrees, D.J. Fox, R.A. Whiteside, R. Seeger, C.F. Melius, J. Baker, R.L. Martin, L.R. Kahn, J.J.P. Stewart, S. Topiol, and J.A. Pople, Gaussian, Inc., Pittsburgh PA, 1990.
5. Ford, G.P. and Scribner, J.D. (1983) *Journal of the American Chemical Society*, 105, 349-354.
6. Traeger, J.C. and McLoughlin, R.G. (1981) *Journal of the American Chemical Society*, 103, 3647-3652.
7. Foster, M.S. and Beauchamp, J.L. (1972) *Journal of the American Chemical Society*, 94, 2425-2431.
8. Foster, M.S., Williamson, A.D., and Beauchamp, J.L. (1974) *International Journal of Mass Spectrometry and Ion Physics*, 15, 429-436.
9. Ford, G.P. (1986) *Journal of the American Chemical Society*, 108, 5104-5108.
10. Stewart, J.J.P. (1988) MOPAC Manual (Fifth Edition) A General Molecular Orbital Package, Air Force Systems Command, United States Air Force.
11. Hehre, W.J., Radom, L., Schleyer, P.v.R., and Pople, J.A. (1986) Ab Initio Molecular Orbital Theory, Wiley & Sons, New York.

12. Saenger, W. (1984) Principles of Nucleic Acid Structure, Springer-Verlag, New York.
13. Pedersen, L.G., Darden, T.A., Deerfield, III, D.W., Anderson, M.W., and Hoel, D.G. (1988) *Carcinogenesis*, 9, 1553-1562.
14. Parthasarathy, R. and Fridey, S.M. (1986) *Carcinogenesis*, 7, 221-227.
15. Clark, T. (1985) A Handbook of Computational Chemistry, Wiley & Sons.
16. Del Bene, (1983) *Journal of Physical Chemistry*, 87, 367.
17. Stewart, J.J.P. (1990) *Journal of Computer-Aided Molecular Design*, 4, 1-105.
18. Hobza, P. and Sandorfy, C. (1987) *Journal of the American Chemical Society*, 109, 1302-1307.
19. Tatewaki, H. and Huzinaga, S. (1979) *Journal of Chemical Physics*, 71, 4339-4348.
20. Schelgel, H.B. (1987) in Ab Initio Methods in Quantum Chemistry - I, Lawley, K.P., ed., Wiley & Sons, pp. 249-285.
21. Dewar, M.J.S., Healy, E.F., and Stewart, J.J.P. (1984) *Journal of the Chemical Society, Faraday Transactions 2*, 80, 227-233.
22. Bartels, R.H. (1972) *University of Texas, Center for Numerical Analysis, Report CNA-44*, Austin, TX.
23. Wolfe, S. and Mitchell, D.J. (1981) *Journal of the American Chemical Society*, 103, 7692-7694.
24. Wolfe, S. and Mitchell, D.J. (1981) *Journal of the American Chemical Society*, 103, 7694-7696.
25. Hammond, G.S. (1955) *Journal of the American Chemical Society*, 77, 334-338.

26. Farneth, W.E. and Brauman, J.I. (1976) *Journal of the American Chemical Society*, 98, 7891-7898.
27. Olmstead, W.N. and Brauman, J.I. (1977) *Journal of the American Chemical Society*, 99, 4219-4228.
28. Stewart, J.J.P. (1989) *Journal of Computational Chemistry*, 10, 209-220.
29. Coolidge, M.B., Marlin, J.E., and Stewart, J.J.P. (1991) *Journal of Computational Chemistry*, 12, 948-952.

CHAPTER 4

Semiempirical Calculations on Guanine Transition States

4.1 Introduction

In this chapter, the results of PM3 calculations concerning the alkylation of guanine are reported. The reactions studied were those between two possible metabolites of N-nitroso compounds, alkyldiazonium ions and carbocations, and two nucleophilic sites in guanine, the O⁶ oxygen and the N⁷ nitrogen, with the N⁹ nitrogen in guanine terminated with a hydrogen. The goal is to calculate the geometries and energies of the transition states involved in these reactions and to infer from the results something about the identity of the ultimate reactive species *in vivo* and the mechanism of the reaction.

The procedures used to optimize the transition states were the same as in the previous section. Coordinate-driving calculations were used to locate an approximate energy maximum between the reactants and products. Geometries were chosen on either side of the maximum and used to find an approximate transition state via the method of Dewar, Healy, and Stewart [1]. The transition state was refined with MOPAC 5.0 using Bartels's method [2] and then vibrationally characterized in a FORCE calculation. Then, these transition state structures were used as starting geometries for MOPAC 6.0 eigenvector following calculations with Baker's

algorithm [3]. The results of these optimizations are presented below.

4.2 Overall Shape of the Potential Surface (MOPAC 5.0)

In the MOPAC 5.0 coordinate driving calculations, the distance between the nucleophilic O⁶ of guanine and the carbon of the incoming alkyldiazonium ion closest to the nitrogens (designated C_m) was decreased incrementally with full minimization of all other coordinates at each fixed C_m-O⁶ distance (6.0 Å, 4.0 Å, 3.0 Å, 2.5 Å, 2.3 Å, 2.1 Å, 1.9 Å, 1.7 Å, 1.5 Å, 1.4 Å, 1.3 Å, and 1.2 Å). This was done for the methyl- ethyl- and propyldiazonium ions and then again, after removing the diazonium nitrogens from the input files, for the reactions involving carbocations. Then, with N⁷ taking the place of O⁶, the process was repeated. The heats of formation at each step is given in Table 4.1 for each type of alkylation (MDO6G = methyldiazonium ion attacking the O⁶ of guanine, ECN7G = ethyl carbocation attacking the N⁷ of guanine, etc.).

One immediate conclusion is apparent. The activation enthalpy barriers for the reactions of the carbocations, if there is any barrier at all, are considerably smaller than those for the reactions of the alkyldiazonium ions. This would be consistent with the view of the carbocation as a fairly indiscriminate S_N1 alkylator. The reactions of the alkyldiazonium ions seem to be characterized by a well-defined ion-molecule complex enthalpy minimum which precedes an enthalpy maximum which is often energetically comparable to the isolated reactants. However, due to the artificiality of this coordinate-driving procedure, little, if any, chemical significance can be attributed to each minimized structure.

Instead, the relationships between optimized maxima and minima, with all coordinates free to vary, are to be analyzed.

Characterization of these points may allow some conclusions to be made concerning the local potential surface.

Table 4.1 - PM3 Heat of Formation Results of Coordinate-Driving Calculations (kcal/mole)

<u>Distance</u>	<u>MDO6G</u>	<u>EDO6G</u>	<u>PDO6G</u>	<u>MDN7G</u>	<u>EDN7G</u>	<u>PDN7G</u>
6.0 Å	225.9	217.6	209.7	221.3	212.1	206.0
4.0 Å	208.9	211.6	204.3	210.5	202.6	196.8
3.0 Å	206.8	199.9	201.5	207.6	201.6	195.5
2.5 Å	211.2	206.3	200.5	213.3	212.4	206.9
2.3 Å	214.3	208.4	202.8	219.6	220.9	215.6
2.1 Å	222.7	218.9	213.4	227.5	200.7	195.1
1.9 Å	214.8	200.6	195.7	197.1	184.9	179.5
1.7 Å	200.5	189.8	184.2	177.4	168.1	162.8
1.5 Å	186.8	180.0	175.0	164.3	157.5	152.2
1.4 Å	185.2	180.1	175.2	165.7	160.2	154.7
1.3 Å	192.7	189.4	184.5	179.4	175.2	169.6
1.2 Å	217.2	215.8	210.8	213.9	211.1	205.2
	<u>MCO6G</u>	<u>ECO6G</u>	<u>PCO6G</u>	<u>MCN7G</u>	<u>ECN7G</u>	<u>PCN7G</u>
6.0 Å	253.4	218.5	211.3	254.2	219.9	213.8
4.0 Å	244.2	197.8	190.4	246.7	201.8	197.1
3.0 Å	235.1	197.2	188.2	237.7	197.0	188.8
2.5 Å	227.5	194.6	187.0	225.8	204.7	196.9
2.3 Å	221.3	195.8	189.3	215.9	195.9	189.8
2.1 Å	211.4	194.2	189.9	200.8	184.9	179.2
1.9 Å	198.0	186.7	185.1	180.9	168.7	163.2
1.7 Å	183.0	173.2	175.7	160.9	151.6	146.2
1.5 Å	170.0	162.7	157.5	147.7	140.8	135.3
1.4 Å	168.3	162.3	157.0	149.0	143.3	137.8
1.3 Å	175.9	171.2	165.8	162.8	158.4	152.7
1.2 Å	201.1	197.6	162.1	197.4	194.5	188.6

A geometry close to the apparent pre-transition state minimum was selected for each reaction and then minimized with respect to all coordinates. Various geometrical parameters for these potential wells are shown in Tables 4.2 and 4.3. N' designates the diazonium nitrogen bound to C_m, and C_e and C_p designate the second and third carbons respectively in the attacking ion.

Comparing the O⁶-C⁶, C⁶-N¹, N⁷-C⁵, and N⁷-C⁸ columns with the PM3 column in Table 3.4 shows that at the ion-molecule complex, the distortion of the bonds near the attack site is small. This is consistent with previous gas-phase *ab initio* (6-31G*) work by Chandrasekhar et. al. where only slight structural changes were apparent between isolated reactants and the ion-molecule complex in the degenerate Cl⁻---CH₃Cl S_N2 reaction [4]. It also seems that the size of the alkyl group is important since the ion-molecule minimum occurs at larger separation for the ethyl and propyl reactions than for the methyl reaction. This may be a steric effect. The relative flatness of the potential surfaces for the carbocation reactions is shown by the fact that three of those minimizations, which were each begun at 3.0 Å separation (all twelve were started at either 3.0 Å or 2.5 Å separation), yielded structures identical to the alkylated products of the reactions. Finally, the incoming C_m carbon seems to maintain a position fairly coplanar with the ring structure (largest deviation for carbocations at O⁶) and, for the O⁶ alkylation, the attack is from the proximal side.

Table 4.2 - PM3 Geometries of the Pre-Transition State Ion-Molecule Complexes at the O⁶ of Guanine (Lengths in Å; Angles in Degrees)

	C _m -O ⁶	C _m -N'	O ⁶ -C ⁶	C ⁶ -N ¹	C _m O ⁶ C ⁶	N ¹ C ⁶ O ⁶ C _m
<u>Reaction</u>	<u>Length</u>	<u>Length</u>	<u>Length</u>	<u>Length</u>	<u>Angle</u>	<u>Dihedral</u>
MDO6G	3.088	1.413	1.223	1.439	107.6	-179.0
EDO6G	3.389	1.446	1.225	1.437	104.0	-176.9
PDO6G	3.239	1.444	1.229	1.431	106.9	171.8
MCO6G	Minimization from 3.0 Å separation yielded product.					
ECO6G	3.682	-	1.232	1.430	112.7	-164.6
PCO6G	3.605	-	1.233	1.429	112.0	-163.1

Table 4.3 - PM3 Geometries of the Pre-Transition State Ion-Molecule Complexes at the N⁷ of Guanine (Lengths in Å; Angles in Degrees)

	C _m -N ⁷	C _m -N'	N ⁷ -C ⁵	N ⁷ -C ⁸	C _m N ⁷ C ⁵	C ⁶ C ⁵ N ⁷ C _m
<u>Reaction</u>	<u>Length</u>	<u>Length</u>	<u>Length</u>	<u>Length</u>	<u>Angle</u>	<u>Dihedral</u>
MDN7G	3.123	1.421	1.404	1.344	100.8	-0.4
EDN7G	3.431	1.457	1.404	1.346	97.6	-5.8
PDN7G	3.430	1.456	1.404	1.346	97.7	-6.5
MCN7G	Minimization from 3.0 Å separation yielded product.					
ECN7G	Minimization from 3.0 Å separation yielded product.					
PCN7G	3.316	-	1.403	1.352	98.3	0.9

There were initially six alkylation products minimized in this study -- O⁶-methylguanine (O6MG), O⁶-ethylguanine (O6EG), O⁶-propylguanine (O6PG), N⁷-methylguanine (N7MG), N⁷-ethylguanine (N7EG), and N⁷-propylguanine (N7PG). Some geometrical parameters are shown in Table 4.4.

Table 4.4 - PM3 Geometries of the Alkylated Products (Lengths in Å; Angles in Degrees)

	C_m-O^6	O^6-C^6	C^6-N^1	N^1-H	$C_mO^6C^6C^5C^6O^6C_m$	
Product	<u>Length</u>	<u>Length</u>	<u>Length</u>	<u>Length</u>	<u>Angle</u>	<u>Dihedral</u>
O6MG	1.423	1.337	1.394	0.999	120.4	0.0
O6EG	1.450	1.329	1.395	1.000	119.3	0.0
O6PG	1.450	1.328	1.395	1.000	119.3	0.3

	C_m-N^7	N^7-C^5	N^7-C^8	$C_mN^7C^5$	$C^6C^5N^7C_m$	
	<u>Length</u>	<u>Length</u>	<u>Length</u>	<u>Angle</u>	<u>Dihedral</u>	
N7MG	1.464	1.407	1.368	125.7	0.9	
N7EG	1.480	1.407	1.368	126.4	1.2	
N7PG	1.479	1.408	1.368	126.2	1.1	

The dihedral angles show that the first carbon of the alkyl group remains coplanar with the rings. Alkylation at N^7 does not distort the adjacent N^7-C^5 or N^7-C^8 bond lengths significantly when compared with Table 3.4. For alkylation at O^6 , the O^6-C^6 and C^6-N^1 bonds have changed significantly from the values in Table 3.4, but the expectation was that the O^6-C^6 bond would become a single bond while the C^6-N^1 bond would become a double bond. When compared with average equilibrium bond lengths of 1.43 Å for a C-O single bond and 1.30 Å for a C=N double bond [5], it is apparent that the expected changes are not complete.

One problem may be that the calculations did not show the loss of the N^1 hydrogen. When this hydrogen was removed from the data file and the O^6 -alkylated products were minimized again, the O^6-C^6 / C^6-N^1 bond lengths became 1.350 Å / 1.349 Å (MDO6G), 1.346 Å /

1.350 Å (EDO6G), and 1.345 Å / 1.350 Å (PDO6G) which are closer to the average values for the expected bond types. However, the loss of the N¹ hydrogen is significantly endothermic, according to the PM3 calculation, unless hydration of the proton is considered. Without hydration, the PM3 heat of reaction for the hydrogen loss is 202.0 kcal/mole, 203.1 kcal/mole, and 203.2 kcal/mole for O6MG, O6EG, and O6PG respectively. If hydration of the proton by three waters is included, by using a MOPAC-minimized H(H₂O)₃⁺ complex in the $\Delta_{\text{rxn}}H$ calculation, the respective heats of reaction decrease to 16.1 kcal/mole, 17.1 kcal/mole, and 17.3 kcal/mole. A bare proton is unlikely in the *in vivo* environment. It seems likely that water is involved in the hydrogen loss, perhaps even as a separate step of the mechanism. The deprotonation of the N¹ nitrogen will be addressed further in Chapter 5.

4.3 Transition States (Bartels's Method - MOPAC 5.0)

Bartels's transition state optimization method as implemented in MOPAC attempts to stepwise reduce the gradient norm. The default number of cycles of geometry optimization is 100 in a single job run. The only one of the transition states here which met the optimization criteria after a single run was the MDO6G transition state. For all the others, in spite of repeated job runs, the gradient was not fully minimized. For the purpose of this study, an optimization was judged complete when the C_m-O⁶(or N⁷) length, the heat of formation and, if applicable, the C_m-N' length changed negligibly for two or three runs with a gradient norm of less than 10

(except for MCO6G where the gradient norm never dropped below 10.488). Some geometrical parameters obtained from these calculations are shown in Tables 4.5 and 4.6.

Table 4.5 - PM3 Geometries of the Transition State (Bartels's Method) for Alkylation at the O⁶ of Guanine (Lengths in Å; Angles in Degrees)

	C _m -O ⁶	C _m -N'	O ⁶ -C ⁶	C ⁶ -N ¹	C _m O ⁶ C ⁶	C ⁵ C ⁶ O ⁶ C _m
<u>Reaction</u>	<u>Length</u>	<u>Length</u>	<u>Length</u>	<u>Length</u>	<u>Angle</u>	<u>Dihedral</u>
MDO6G	2.064	1.912	1.252	1.415	118.4	0.1
EDO6G	2.294	2.207	1.245	1.418	116.4	10.5
PDO6G	2.364	2.233	1.242	1.420	116.5	9.1
MCO6G	3.481	-	1.229	1.433	142.2	1.5
ECO6G	2.624	-	1.237	1.425	121.8	-12.6
PCO6G	2.697	-	1.237	1.426	118.4	12.4

Table 4.6 - PM3 Geometries of the Transition State (Bartels' Method) for Alkylation at the N⁷ of Guanine (Lengths in Å; Angles in Degrees)

	C _m -N ⁷	C _m -N'	N ⁷ -C ⁵	N ⁷ -C ⁸	C _m N ⁷ C ⁵	C ⁴ C ⁵ N ⁷ C _m
<u>Reaction</u>	<u>Length</u>	<u>Length</u>	<u>Length</u>	<u>Length</u>	<u>Angle</u>	<u>Dihedral</u>
MDN7G	2.138	1.801	1.404	1.352	112.8	175.6
EDN7G	2.396	2.148	1.405	1.351	110.0	179.0
PDN7G	2.549	2.186	1.405	1.350	108.4	177.8
MCN7G	2.498	-	1.401	1.348	100.0	-179.5
ECN7G	2.499	-	1.408	1.356	127.9	169.5
PCN7G	2.349	-	1.408	1.359	117.9	164.5

Conventional wisdom and x-ray crystallography [6] calls for the O⁶-C⁶ double bond to become a single bond while the C⁶-N¹ single bond becomes a double bond as the N¹ hydrogen is forced off.

While the N¹-H bond length is essentially constant throughout these simulations, Table 4.5 does show some lengthening of the O⁶-C⁶ bond and some shortening of the C⁶-N¹ bond at the transition state. A steric effect is apparent for the diazonium ion reactions as the C_m-O⁶ bond lengths increase with increasing alkyl group size. This trend is not present in the carbocation reactions. The MCO6G transition state has an especially long C_m-O⁶ length. However, as will be discussed later, this may be due to the relative flatness of the potential surfaces for the carbocation reactions.

To really compare these reaction mechanisms with experimental observations, it is necessary to look at the energetics of the calculations. Table 4.7 lists the heats of reaction, activation enthalpies ($=\Delta_f H_{\text{transition state}} - \Delta_f H_{\text{reactants}}$), and intrinsic barriers ($=\Delta_f H_{\text{transition state}} - \Delta_f H_{\text{ion-molecule complex}}$) for each reaction based on the structures optimized with Bartels's method. In general, the reactions of the carbocations are more exothermic than those of the alkyldiazonium ions. This, considering the significantly negative activation enthalpies shown in Table 4.7 and the apparent lack of reaction barriers in the coordinate-driving calculations, is consistent with the carbocation being an indiscriminately reactive species. There is no clear indication why there would be a difference in alkylation at O⁶ versus N⁷.

For alkylation by alkyldiazonium ion, the picture is different. Alkylation at the N⁷ position is consistently more exothermic than at the O⁶ position. The activation enthalpies for both sites are comparable for ethylation and propylation, but is somewhat less for

methylation at N⁷ than for methylation at O⁶. The intrinsic barriers are also lower for alkylation at N⁷ than at O⁶, but, for ethylation, the difference is very small. These relationships are consistent with the experimental observations that the more likely alkylation site is the N⁷ position, and that the preference for N⁷ over O⁶ is less pronounced for ethylation than for methylation (see Conclusions).

Table 4.7 - PM3 Heats of Reaction, Activation Enthalpies and Intrinsic Barriers (kcal/mole)

<u>Reaction</u>	<u>Heat of Reaction</u>	<u>Activation Enthalpy</u>	<u>Intrinsic Barrier</u>
MDO6G	-48.0	-2.8	24.1
EDO6G	-44.9	-6.6	19.1
PDO6G	-44.0	-7.4	23.1
MDN7G	-68.9	-5.6	20.5
EDN7G	-66.5	-6.8	18.8
PDN7G	-66.2	-6.8	19.6
MCO6G	-96.0	-20.8	-
ECO6G	-67.8	-26.0	9.0
PCO6G	-64.6	-24.6	10.9
MCN7G	-116.9	-35.4	-
ECN7G	-89.4	-25.6	-
PCN7G	-86.8	-26.0	10.6

No intrinsic barriers are listed for MCO6G, MCN7G, or ECN7G because of the problems with optimizing their ion-molecule complexes (see Tables 4.2 and 4.3 and following discussion). The problems with these optimizations may be the relatively smoothly decreasing energy profiles of the reactions of the carbocations. This is also seen by looking at the force constants for the structures optimized by Bartels's method -- that is, the structures which have

been discussed as transition states thus far. Of course, a genuine transition state has exactly one negative force constant. Table 4.8 shows the two algebraically smallest force constants for each "transition state" optimized with Bartels's method. All other force constants are positive.

Table 4.8 - PM3 Force Constants for Transition States Optimized by Bartels's Method (millidynes/Å)

<u>Reaction</u>	<u>Force Constant 1</u>	<u>Force Constant 2</u>
MDO6G	-1.2931	0.0042
EDO6G	-0.3442	-0.0009
PDO6G	-0.2330	-0.0062
MDN7G	-1.5513	0.0053
EDN7G	-0.4747	-0.0011
PDN7G	-0.2899	-0.0031
MCO6G	-0.0057	-0.0004
ECO6G	-0.0203	-0.0083
PCO6G	-0.0012	-0.0003
MCN7G	-0.2899	0.0033
ECN7G	-0.2329	-0.0072
PCN7G	-0.4290	0.0007

For all but three of the structures, namely, MCO6G, ECO6G, and PCO6G, the force constant analysis is such that the structures can be considered transition states. The negative values of force constant 2 for EDO6G, PDO6G, EDN7G, PDN7G, and ECN7G are so close to zero as to be negligible. Following that practice, however, the structures optimized for the reaction of the carbocations at the O⁶ position appear to be minima instead of transition states. Of course, if there is a pre-product minimum for an exothermic reaction, then, logically, there must be a maximum between that

Table 4.9 - PM3 Force Constants for Typical S_N2 Transition States Optimized by Bartels's Method (millidynes/Å)

<u>Transition State</u>	<u>Force Constant 1</u>	<u>Force Constant 2</u>
Cl--CH ₃ --Cl	-1.0342	0.1217
HCC--CH ₃ --CCH	-2.5646	0.0090
NC--CH ₃ --F	-3.4012	0.0563
F--CH ₃ --F	-3.4612	0.1858
CN--CH ₃ --OH	-1.8277	-0.0020
HO--CH ₃ --F	-3.1735	0.0146
HCC--CH ₃ --F	-3.5181	0.0322
HO--CH ₃ --OH	-1.9490	0.0121
H ₂ N--CH ₃ --F	-2.8804	0.0043
CH ₃ O--CH ₃ --F	-3.2144	0.0016
CH ₃ O--CH ₃ --OH	-2.0817	-0.0007

minimum and the product. Bartels's method will find either a transition state or some kind of minimum depending on the nature of the potential surface and the starting geometry [7]. The apparent failure of Bartels's method to find a genuine transition state in these three cases may reflect the complexity and, again, the flatness of the potential surfaces. Table 4.9 lists the MOPAC force constants for some of the transition states discussed in the previous section. Comparison of these with Table 4.8 show that the force constants for the guanine alkylation transition states are smaller. This and the near-zero values of force constant 2 indicates that these transition states are less clearly-defined and more loosely-bound than typical S_N2 transition states. This is consistent

with the suggestion by Scribner and Ford [8] that the guanine transition states are loose enough to allow rearrangement of the propyl ion at the transition state to form the isopropyl adduct.

4.4 Eigenvector Following Calculations (MOPAC 6.0)

In this section, the results of calculations similar to those in sections 4.2 and 4.3 are reported except that, here, the Baker eigenvector following routine [3] in MOPAC 6.0 was used for both energy minimizations and transition state optimizations. Tables 4.10 - 4.14 are of the same format as, and, therefore, analogous to Tables 4.4 - 4.8. Since all calculations in this chapter were performed with the PM3 method, the results in Tables 4.10 - 4.14 should, in principle, agree with the results in Tables 4.4 - 4.8. After all, it is the PM3 Hamiltonian form which determines the energy calculated for a given geometry. The only difference in these two sets of calculations is in the optimization algorithm - that is, in the method of exploring the PM3 potential surface. However, this sometimes makes a significant difference, especially in the optimization of a transition state.

Comparing Table 4.10 with Table 4.4 shows very little difference between the MOPAC 5.0 results and the MOPAC 6.0 results. Despite preliminary indications of increased optimization speed with the eigenvector following routine [9], the computation time required for the MOPAC 6.0 calculations using Baker's algorithm were usually comparable to those of the MOPAC 5.0 calculations for the optimizations of Tables 4.4 and 4.10. In fact, there were two cases, namely N7MG and N7PG, for which the MOPAC

5.0 computation time was significantly longer. However, this was not really a fair test of the two algorithms. The two sets of calculations shared common starting geometries which happened to be the previously-optimized MOPAC 5.0 structures. Baker's algorithm may, indeed, be faster than MOPAC 5.0's default algorithm if each is started from a common point significantly displaced from the optimum geometry.

**Table 4.10 - PM3 Geometries of the Alkylated Products (Eigenvector Following)
(Lengths in Å; Angles in Degrees)**

	C_m-O^6	O^6-C^6	C^6-N^1	N^1-H	$C_mO^6C^6$	$C^5C^6O^6C_m$
<u>Product</u>	<u>Length</u>	<u>Length</u>	<u>Length</u>	<u>Length</u>	<u>Angle</u>	<u>Dihedral</u>
O6MG	1.423	1.337	1.394	0.999	119.5	0.0
O6EG	1.450	1.329	1.395	1.000	119.3	0.0
O6PG	1.450	1.328	1.395	1.000	119.3	0.3
	C_m-N^7	N^7-C^5	N^7-C^8	$C_mN^7C^5$	$C^6C^5N^7C_m$	
	<u>Length</u>	<u>Length</u>	<u>Length</u>	<u>Angle</u>	<u>Dihedral</u>	
N7MG	1.464	1.407	1.368	125.7	-0.5	
N7EG	1.480	1.407	1.368	126.4	1.1	
N7PG	1.479	1.407	1.368	126.2	1.0	

Comparing Tables 4.11 and 4.12 with Tables 4.5 and 4.6 show some similarities and some striking differences between the MOPAC 5.0 and MOPAC 6.0 results. The largest differences occur for the transition states of EDO6G, PDO6G, MCO6G, EDN7G, PDN7G, and ECN7G. Examination of the force constants given in Table 4.14 which were calculated at the "Baker-optimized" transition state

Table 4.11 - PM3 Geometries of the Transition State (Eigenvector Following) for Alkylation at the O⁶ of Guanine (Lengths in Å; Angles in Degrees)

	C _m -O ⁶	C _m -N'	O ⁶ -C ⁶	C ⁶ -N ¹	C _m O ⁶ C ⁶	C ⁵ C ⁶ O ⁶ C _m
<u>Reaction</u>	<u>Length</u>	<u>Length</u>	<u>Length</u>	<u>Length</u>	<u>Angle</u>	<u>Dihedral</u>
MDO6G	2.064	1.909	1.252	1.415	118.4	0.6
EDO6G	3.191	2.200	1.236	1.427	104.2	5.3
PDO6G	3.219	2.151	1.236	1.427	104.4	5.7
MCO6G	2.377	-	1.241	1.417	111.2	0.6
ECO6G	2.854	-	1.236	1.426	110.7	-3.9
PCO6G	2.864	-	1.236	1.426	108.6	5.8

Table 4.12 - PM3 Geometries of the Transition State (Eigenvector Following) for Alkylation at the N⁷ of Guanine (Lengths in Å; Angles in Degrees)

	C _m -N ⁷	C _m -N'	N ⁷ -C ⁵	N ⁷ -C ⁸	C _m N ⁷ C ⁵	C ⁴ C ⁵ N ⁷ C _m
<u>Reaction</u>	<u>Length</u>	<u>Length</u>	<u>Length</u>	<u>Length</u>	<u>Angle</u>	<u>Dihedral</u>
MDN7G	2.136	1.787	1.404	1.352	113.6	179.8
EDN7G	3.304	2.084	1.405	1.350	97.2	172.1
PDN7G	3.290	2.028	1.404	1.350	98.2	171.8
MCN7G	2.522	-	1.401	1.348	99.9	179.9
ECN7G	3.088	-	1.402	1.348	108.5	166.4
PCN7G	2.354	-	1.409	1.359	117.2	166.4

geometries shows that each structure has exactly one negative force constant as required of a true transition state. This fact may lend a greater confidence in the results in Tables 4.11 and 4.12 *versus* Tables 4.5 and 4.6. It should be pointed out, however, that, similar to the results in Table 4.8, Table 4.14 gives force constants for the ECO6G and PCO6G transition states which are very close to zero so that these structures may instead be shallow minima, or, at least, points on a fairly flat energy surface. In contrast, since Force

Constant 1 in Table 4.14 for the MCO6G structure is significantly negative, the Baker-optimized geometry can confidently be called a transition state. This was not the case in the MOPAC 5.0 result of Table 4.8.

Examining the results of the enthalpy calculations in Table 4.13 shows trends similar to those of Table 4.7. Alkylation by carbocation is again more exothermic and is associated with lower activation enthalpies than alkylation by diazonium ion. It is interesting to note that no satisfactory intrinsic barriers could be calculated for the carbocation reactions. In the cases of MCO6G, MCN7G, and ECN7G, the problem was similar to that of the MOPAC 5.0 calculations in that minimization from close to a probable ion-molecule energy well (using starting geometries from MOPAC 5.0 calculations) led to products. In the other cases, the eigenvector following routine located what appeared to be a reasonable ion-molecule complex, but warned that the structure did not correspond with a stationary point on the energy surface.

In the diazonium ion calculations, the observed preference for N⁷ alkylation *versus* O⁶ alkylation is predicted by the more negative $\Delta_{rxn}H$, lower activation enthalpies, and lower intrinsic barriers of the N⁷ reactions. In the next section, the conclusions drawn from these semiempirical calculations are discussed.

Table 4.13 - PM3 Heats of Reaction, Activation Enthalpies and Intrinsic Barriers - Eigenvector Following (kcal/mole)

<u>Reaction</u>	<u>Heat of Reaction</u>	<u>Activation Enthalpy</u>	<u>Intrinsic Barrier</u>
MDO6G	-48.1	-2.8	27.8
EDO6G	-44.9	-11.8	13.9
PDO6G	-44.0	-13.2	17.3
MDN7G	-68.9	-6.1	23.6
EDN7G	-66.5	-15.0	10.5
PDN7G	-66.2	-16.7	9.7
MCO6G	-96.0	-35.4	-
ECO6G	-67.8	-30.1	-
PCO6G	-64.6	-29.2	-
MCN7G	-116.9	-35.4	-
ECN7G	-89.4	-32.5	-
PCN7G	-86.8	-26.1	-

Table 4.14 - PM3 Force Constants for Transition States Optimized by Eigenvector Following (millidynes/Å)

<u>Reaction</u>	<u>Force Constant 1</u>	<u>Force Constant 2</u>
MDO6G	-1.2957	0.0042
EDO6G	-0.1688	0.0018
PDO6G	-0.2008	0.0016
MDN7G	-1.5760	0.0051
EDN7G	-0.2052	0.0032
PDN7G	-0.2460	0.0016
MCO6G	-0.2698	0.0030
ECO6G	-0.0044	0.0023
PCO6G	-0.0049	0.0013
MCN7G	-0.2698	0.0030
ECN7G	-0.1334	0.0028
PCN7G	-0.4017	0.0009

4.5 Conclusions

The goal of this study was to characterize the alkylation of guanine by N-nitroso compounds so that some conclusions could be made about the ultimate reactive species of these mutagens and so that some inferences could be drawn concerning the mechanism of the alkylation. The reactive species investigated were the methyl-ethyl- and propyldiazonium ions and the corresponding carbocations.

The results reported in this study indicate that the more likely ultimate mutagen is the alkylidiazonium ion. There are several experimental observations which must be accounted for in reaching this conclusion. They are: 1) The alkylation of the N⁷ position of guanine is favored over alkylation at the O⁶ position [10], 2) While alkylation occurs mainly at nitrogen centers on guanine, the relative reactivity toward oxygen sites is greater for ethylating agents than for methylating agents [11,12], and 3) In general, methylating agents are more reactive with nucleic acids than are ethylating agents which, in turn, are more reactive than propylating agents [13].

The overall shapes of the potential surfaces for the carbocation reactions are inconsistent with these experimental observations. The carbocation energy surfaces are considerably smoother than those of the alkylidiazonium ion reactions. This is shown in the way the heat of formation varies with ion-molecule separation in Table 4.1, in the difficulty in finding a stable ion-molecule complex for several of the carbocation reactions, in the heats of reactions and activation enthalpies in Tables 4.7 and 4.13, and in the value of the force constants in Tables 4.8 and 4.14

for the optimized transition states. These factors were evident in both the MOPAC 5.0 calculations and in the eigenvector following calculations using MOPAC 6.0. A smoothly decreasing reaction profile, as is apparent in the carbocation reactions, is not consistent with any preference for one nucleophilic site in guanine over another.

On the other hand, the calculated enthalpy parameters in Tables 4.7 and 4.13 for alkylation by diazonium ion seem to favor the alkyldiazonium ion reactions as a more satisfactory explanation of the experimental observations. In both sets of calculations, alkylation at N⁷ is associated with more negative heats of reaction, lower activation enthalpies, and smaller intrinsic barriers when compared to alkylation at O⁶. This is perhaps more pronounced in the MOPAC 6.0 results. These trends are consistent with the preference of alkylating agents to alkylate N⁷ *versus* O⁶. The MOPAC 5.0 activation enthalpies for methylation and ethylation at O⁶ *versus* N⁷ predict, as is observed in experiments, that the preference to alkylate at N⁷ is less pronounced for ethylation than for methylation. The higher reactivity of ethylating agents over propylating agents is correctly predicted by the lower intrinsic barriers for ethylation. As mentioned in Chapter 3, these intrinsic barriers are useful indicators of reaction efficiency when the transition state energy is lower than that of the reactants [14]. Unfortunately, the intrinsic barriers for methylation are calculated to be higher than for either ethylation or propylation. However, in view of the consistently smaller C_m-O⁶(or N⁷) separations for methylation, both at the ion-molecule complex and at the transition

state, it may be that the greater reactivity for methylating agents is due to a steric effect not apparent in these calculations on a single guanine base unit.

This highlights one of the caveats which must be placed on conclusions from these calculations. These were gas-phase calculations on a single guanine. The influence of the rest of the DNA molecule and of the surrounding aqueous media may be significant. There is some reassurance from the work of Milligan *et. al.* on methylation by MNU. They found the same N⁷ methylation patterns for reactions of MNU with both B and non-B forms of DNA and, therefore, suggest that electronic factors dominate the reaction [15]. However, some influence of neighboring residues is apparent in the increased likelihood of guanine alkylation when the guanine is preceded 5' by another purine residue [16].

Still, the results in this chapter are significant in that they represent the first semiempirical transition state calculations for the alkylation of guanine by N-nitroso compounds which compare the kinetic barriers for alkylation by alkyldiazonium ions and carbocations. Ford and Scribner had achieved analogous results for the relative reactivities of methyl- and ethyldiazonium ions with small molecules representing various nucleophilic sites in nucleic acid bases [17]. Miertus and Trebaticka had studied the equilibrium thermodynamics of the reactions of methyl and ethyl carbocations with nucleic acid bases [18]. This study has combined and extended the concepts in these previous works by performing comparative calculations on both alkyldiazonium ions and carbocations in their reactions with guanine using MOPAC's newest method, PM3. The

apparent agreement of PM3 with experimental observations lends credibility to semiempirical transition state calculations in general, and to the PM3 method, in particular.

However, these calculations are unsatisfactory, or, at least, incomplete, in that they do not account for the loss of the N¹ hydrogen. There is specific x-ray crystallographic evidence of the lack of a hydrogen on the N¹ nitrogen after methylation at O⁶, both in O⁶-methylguanosine [19] and in O⁶-methylguanine as part of a double-strand segment [20]. However, the calculations in this chapter indicate no lengthening of the N¹-H bond upon alkylation. Either the semiempirical methods used were unable to effectively represent the bond-breaking and bond-forming phenomena in the reactions, or other factors, such as the presence of water, help to effect the deprotonation of the N¹ nitrogen. In the next chapter, this matter is investigated further.

References

1. Dewar, M.J.S., Healy, E.F., and Stewart, J.J.P. (1984) *Journal of the Chemical Society, Faraday Transactions 2*, 80, 227-233.
2. Bartels, R.H. (1972) *University of Texas, Center for Numerical Analysis, Report CNA-44*, Austin, TX.
3. Baker J. (1986) *Journal of Computational Chemistry*, 7, 385-395.
4. Chandrasekhar, J., Smith, S.F., and Jorgensen, W.L., *Journal of the American Chemical Society*, 107 (1985) 154-163.
5. Weinberger, S. J. and Stermitz, F.R., Organic Chemistry, Academic Press, 1984, p.54.
6. Parthasarthy, R. and Fridey, S.M. (1986) *Carcinogenesis*, 7, 221-227.
7. Stewart, J.J.P. (1988) MOPAC Manual (Fifth Edition) A General Molecular Orbital Package, Air Force Systems Command, United States Air Force.
8. Scribner, J.D. and Ford, G.P. (1982) *Cancer Letters*, 16, 51-56.
9. Stewart, J.J.P. (1990) MOPAC Manual (Sixth Edition) A General Molecular Orbital Package, Air Force Systems Command, United States Air Force.
10. Lawley, P.D. (1989) *Mutation Research*, 213, 3-25.
11. Singer, B. (1976) *Nature*, 264, 333-339.
12. Singer, B. (1979) *Journal of the National Cancer Institute*, 62, 1329-1339.
13. Singer, B. and Kusmierck, J.T. (1982) *Annual Review of Biochemistry*, 51, 655-693.

14. Olmstead, W.N. and Brauman, J.I. (1977) *Journal of the American Chemical Society*, 99, 4219-4228.
15. Milligan, J.R., Snotnicki, S.M., and Archer, M.C., *Carcinogenesis*, 11 (1990) 947-951.
16. Horsfall, M.J., Gordon, A.J.E., Burns, P.A., Zielenska, M., van der Vliet, G.M.E., and Glickman, B.W. (1990) *Environmental and Molecular Mutagenesis*, 15, 107-122.
17. Ford, G.P. and Scribner, J.D. (1983) *Journal of the American Chemical Society*, 105, 349-354.
18. Miertus, S. and Trebaticka, M. (1984) *Journal of Theoretical Biology*, 108, 509-517.
19. Parthasarathy, R. and Fridley, S.M. (1986) *Carcinogenesis*, 7, 221-227.
20. Leonard, G.A., Thomson, J., Watson, W.P., and Brown, T. (1990) *Proceedings of the National Academy of Sciences, USA*, 87, 9573-9576.

CHAPTER 5

The Deprotonation of the N¹ Nitrogen

5.1 Introduction and Methods

The principle conclusion made in the preceding chapter was that the probable ultimate DNA-alkylating metabolite of N-nitroso compounds is an alkyldiazonium ion. However, as pointed out in section 4.5, the semiempirical analysis of the reactions of alkyldiazonium ions with guanine did not account for the loss of the proton at the N¹ position upon O⁶ alkylation. The importance of this hydrogen loss in the mutagenic mechanism dates back to the suggestion by Loveless [1] that O⁶-alkylation of guanine and subsequent N¹ deprotonation could lead to anomalous base pairing of O⁶-alkylguanine with thymine which could effect the G:C → A:T transition (see Figures 1.3 and 1.4). An x-ray crystallographic structure determination by Parathasarthy and Fridey [2] confirmed the lack of the N¹ hydrogen in O⁶-methylguanosine. The anomalous base pairing suggested by Loveless has been characterized in an x-ray structure study of a self-complementary dodecanucleotide containing two O⁶methylG:T base pairs [3].

Since, from results in Chapter 4, abstraction of the free proton at N¹ was found to be significantly endothermic, and since including water in the calculation made the process less endothermic (see end of section 4.2), it seems reasonable to further investigate the

possible role of water in the deprotonation of the N¹ nitrogen. In this chapter, this concept is explored further using both semiempirical and density-functional techniques. All MOPAC calculations reported in this chapter were performed with the PM3 method. Most were done with MOPAC 5.0 [4]. The exceptions, where MOPAC 6.0 [5] was used, were the calculation of atomic charges fitted to an electrostatic potential (to be discussed in the next section) and the MDO6G transition state optimization with a water molecule included (in section 5.3). All density-functional calculations were performed using the UniChem implementation of DGauss [6].

In the DGauss calculations, the Becke-Perdew correction to the local density approximation (see section 2.5) was used [6-10]. The local density approximation can yield reasonable molecular geometries, but the energetics of chemical reactions are not well predicted [6]. In the Becke-Perdew correction, the determination of the exchange-correlation energy includes the local values of the derivatives of the electron density instead of just the local density alone. The correction is "added on" to the end of the DGauss calculation. That is, the initial SCF calculations and subsequent geometry optimizations are accomplished using the local density approximation. Then, the Becke-Perdew correction is calculated at the final geometry obtained within the local density approximation [6].

DGauss offers several basis set options each of which has been optimized for use in with the local density approximation [6]. In this study, unless noted otherwise, the orbital and auxiliary basis

sets used (see section 2.5) were DGauss's DZVP2 and A2 basis sets respectively. The DZVP2 set is the better of two available double-zeta-split-valence+polarization basis sets. For carbon, oxygen, and nitrogen the basis, denoted (721/51/1)/[3/2/1], consists of three s-type functions (one comprising seven primitive Gaussians, one comprising two primitive Gaussians, and one uncontracted Gaussian), two p-type functions (one comprising five primitive Gaussians and one uncontracted), and one d-type uncontracted polarization function. For hydrogen, the DZVP2 basis is denoted (41/1)/[2/1] and is made up of three s-type functions (one comprising four primitive Gaussians and one uncontracted) and a single p-type polarization function [6]. Thus, the DZVP2 basis is comparable to the 6-31G** of Gaussian 90 [11]. The A2 auxiliary basis set is of the form [4/1] for hydrogen and [8/4/4] for carbon, oxygen, and nitrogen. The auxiliary basis sets for the electron density and for the exchange-correlation energy (see page 38) are similar except for the values of the Gaussian exponents [6].

For the MOPAC calculations, the energy minimization and transition state optimization routines were discussed in Chapter 2. In DGauss, geometry optimizations are accomplished using analytic energy gradients in a quasi-Newton search algorithm with Hessian updating using the BFGS method [6,12]. Transition state optimization in DGauss proceeds through minimization of the gradient norm by the method of McIver and Komornicki which requires a reasonable Hessian matrix be provided at the beginning of the calculation [6,13]. In the DGauss calculations, all convergence and optimization criteria were set to the default "medium" option

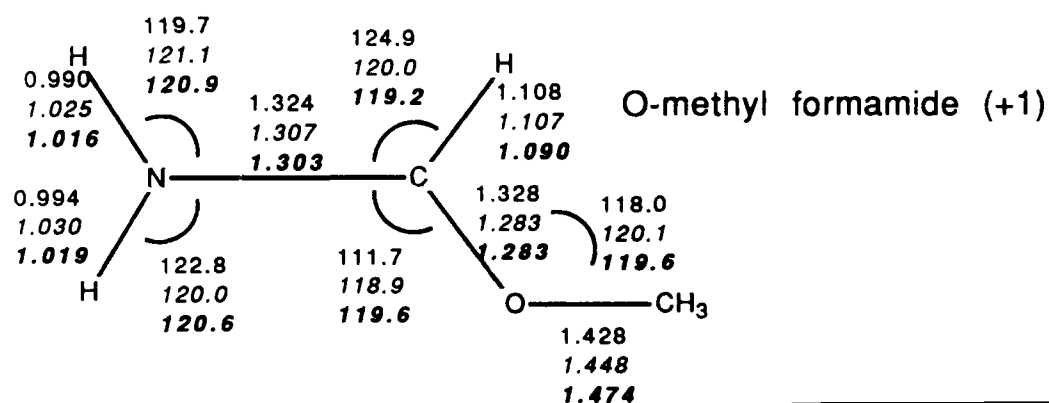
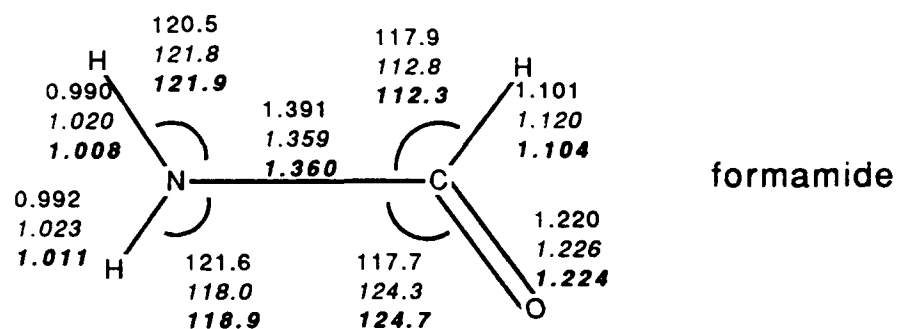
unless noted otherwise. Two sample DGauss input files are shown in the Appendix.

5.2 The Effect of Water on Formamide-Based Models

Following, as in Chapter 3, the example of Ford and Scribner [14], formamide was chosen as a model of the O⁶ site in guanine for the purpose of this study's preliminary calculations on the involvement of water in the deprotonation of guanine's N¹ nitrogen upon O⁶ alkylation. The use of this model was necessary to allow sufficiently high level traditional *ab initio* (Gaussian 90 [15]) calculations to be performed to provide a basis for comparison with MOPAC and DGauss results.

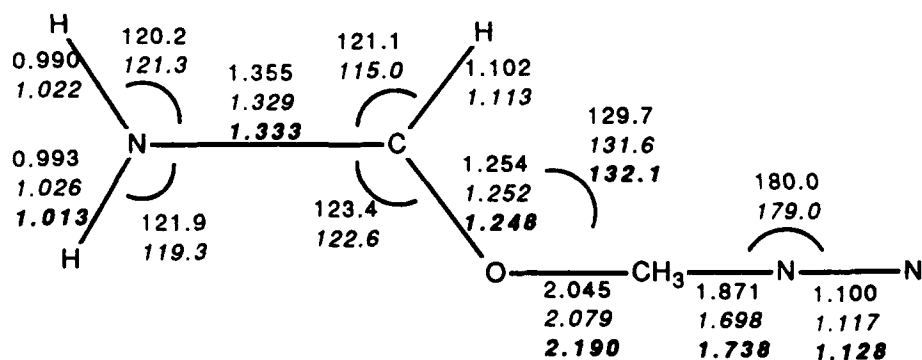
The first step was to perform DGauss calculations on the formamide-based systems reported in Chapter 3 - namely, formamide, positively-charged O-methyl formamide, and the transition state for the methylation of formamide by the methyldiazonium ion. The necessary input Hessian for the DGauss transition state optimization was produced by a single-point DGauss calculation of the energy second derivatives at the input geometry. This technique was used for all DGauss transition state optimizations except for the MDO6G...H₂O transition state in section 5.3 where the initial Hessian was produced by a similar single-point calculation using MNDO90 [6]. The results of these optimizations are shown in Figure 5.1 along with the MOPAC 5.0 PM3 and Gaussian 90 MP2/6-31G* results from Chapter 3. The MP2/6-31G* transition state calculation was not a full geometry optimization (see Chapter 3) and the parameters shown are the optimized active variables.

Figure 5.1 - Optimized Geometries of
Formamide-Based Systems



transition state for the
methylation of formamide
by the methyldiazonium ion (+1)

Legend:
PM3
DGauss
MP2/6-31G*
(if applicable,
see text)



When compared with the other two sets of calculations, DGauss's density-functional results are entirely reasonable. Where there are significant differences, the DGauss and MP2/6-31G* results tend to agree more closely with each other than with the MOPAC results. This is most apparent in the distance between the nitrogen and carbon of formamide and, to a lesser extent, in this same bond in the formamide derivatives. The DGauss prediction of the O-C(methyl) bond length in the transition state and in the methylated formamide is intermediate between the MOPAC and Gaussian 90 results but somewhat closer to the MOPAC result. Most significant for this chapter, however, is that, similar to the other two sets of calculations, the DGauss results show no significant increase in the N-H bond length either at the transition state or in the methylated formamide.

The favorable comparison between DGauss and MP2/6-31G* indicates that DGauss may be used with reasonable confidence for these systems. Since transition states are usually harder to optimize than are equilibrium structures, a separate DGauss transition state optimization was performed for the degenerate S_N2 reaction $Cl^- + CH_3Cl \rightarrow ClCH_3 + Cl^-$. Even using the DZVP orbital basis set and the A1 auxiliary basis set (more restricted choices than the DZVP2 and A2 sets used elsewhere in this study) the DGauss optimized C-Cl distance at the transition state was 2.303 Å which compares very well with the MP2/6-31G* value of 2.308 Å. More evidence of the predictive capability of DGauss will be presented in the next section where comparisons are made between DGauss-optimized and experimental bond lengths for

guanine and O⁶-methylguanine. As the influence of water on formamide-based systems is now discussed, MOPAC PM3 and DGauss density-functional results will be presented for comparison.

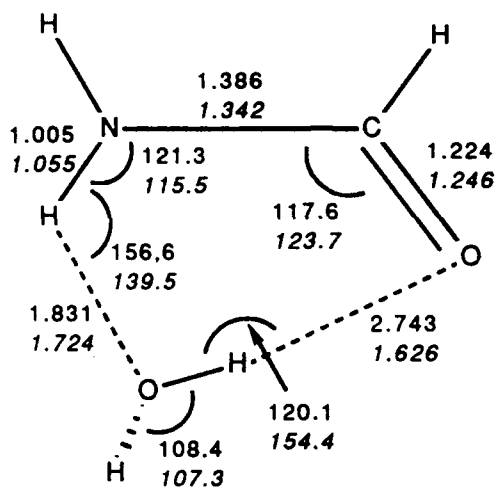
In each structure shown in Figure 5.1, a water molecule was then placed so that the oxygen atom of the water was 1.5 Å from the hydrogen designated H¹ in Figure 3.4. The water molecule was oriented so that it was coplanar with the formamide system with the water oxygens pointing away from the H¹ hydrogen. These new structures were then optimized using PM3 (MOPAC 5.0 - Bartels's method for the transition state [16]) and DGauss. The results of these optimizations are shown in Figures 5.2 and 5.3. All calculations allowed full geometry optimization.

The most obvious difference among the structures in Figures 5.2 and 5.3 is in the optimized placement and orientation of the water molecule. In the "formamide with one water" structure, the position of the water molecule is significantly affected by hydrogen bonding to the formamide oxygen. The effect is most apparent in the DGauss results where the two hydrogen-bonding distances are comparable. One of the water's hydrogens remains coplanar with the formamide while the other is forced out of the plane by about 40° to 70° depending on the method. At the same time, there was some small lengthening of the C-O bond, some shortening of the C-N bond, and some lengthening of the N-H¹ bond due to the interaction with the water, but the nitrogen still has not been deprotonated.

In the optimized "O-methyl formamide (+1) with water" structure in Figure 5.2 and the transition state in Figure 5.3, the position of the water molecule is relatively unaffected by the

Figure 5.2 - The Effect of Water on Formamide
and O-methyl Formamide (+1)

formamide with one water



Legend:
PM3 (MOPAC 5.0)
DGauss

O-methyl formamide (+1) with one water

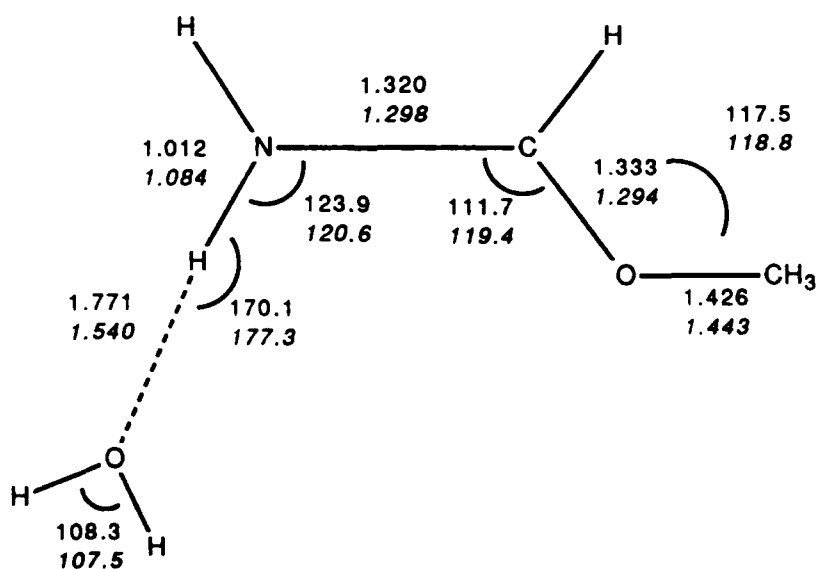
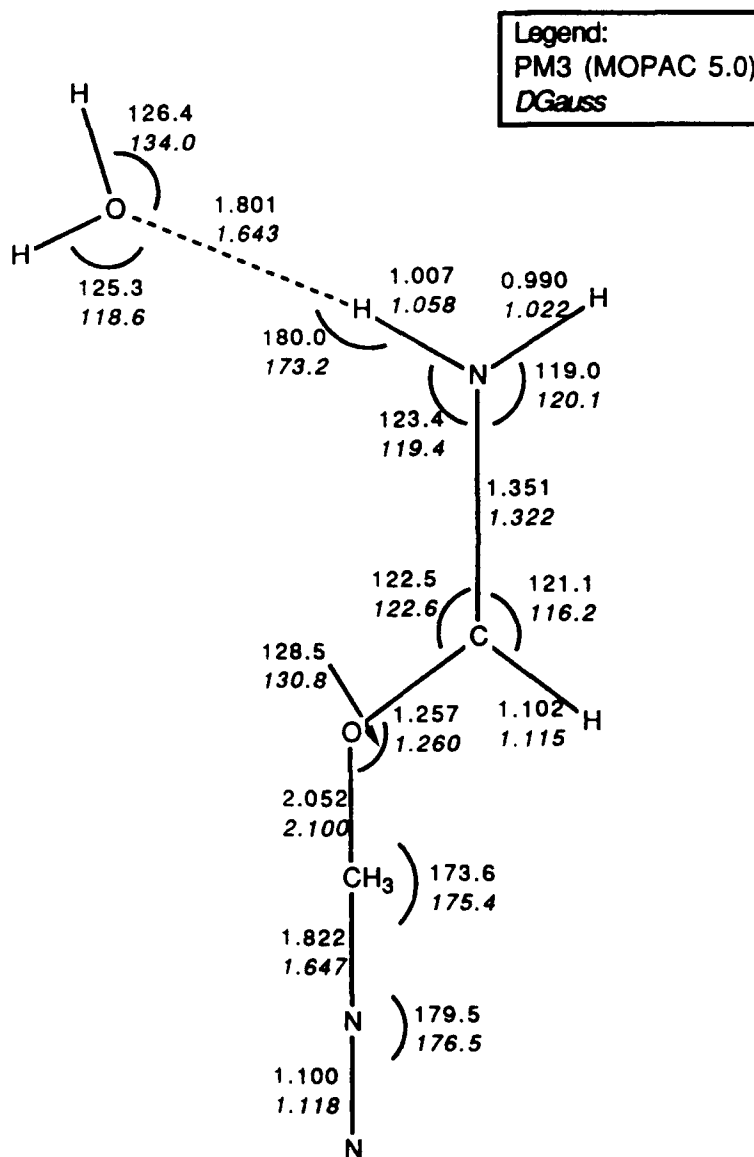


Figure 5.3 - The Effect of Water on the Transition State for the Methylation of Formamide by the Methyldiazonium Ion



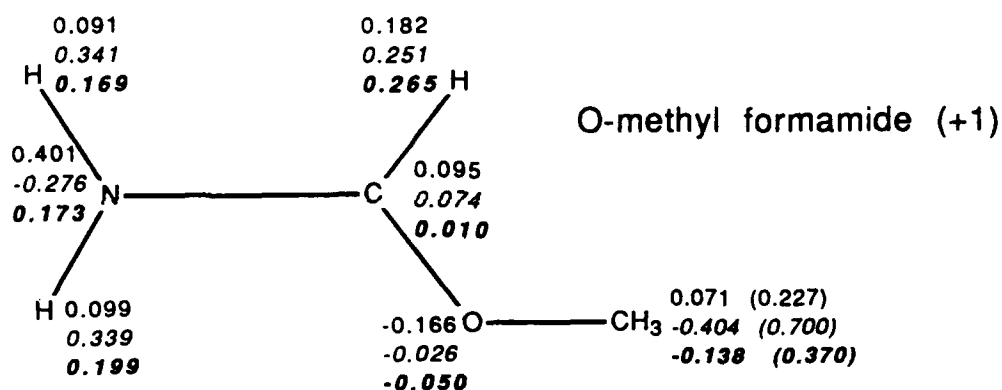
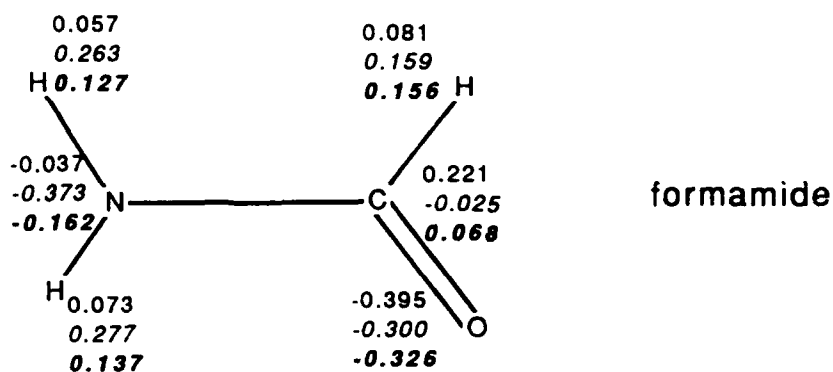
formamide oxygen. There are some changes in internal bond lengths, but no deprotonation of the nitrogen. The reason for the differences in the structures in Figures 5.2 and 5.3 may be explained by reference to the calculated atomic charges shown in Figure 5.4. In Figure 5.4 are shown charges calculated by the standard PM3 method of MOPAC 5.0, the Mulliken charge analysis of DGauss, and the charges fitted to an electrostatic potential (ESP) calculated by MOPAC 6.0 [5] at PM3 eigenvector-following optimized geometries [17] using the method of Besler, Merz, and Kollman [18]. For methyl groups, the charge on the carbon is given first and, then, the sum of the charges on the hydrogens is given in parentheses.

Both the MOPAC 5.0 and DGauss atomic charges are based on a Mulliken population analysis. This type of approach has been necessary because atomic charges are not directly obtainable from an electronic wave function. The analysis is based on an arbitrary partitioning of the overlap electron population between pairs of atomic orbital basis functions such that each orbital is assigned half of the overlap population [19]. Charges thus assigned are basis set dependent and, therefore, not always reliable. However, unlike atomic charges, the electrostatic potential as a function of position may be directly calculated from the wave function [20]. The electrostatic potential, $V(r)$ is given by [18]

$$V(r) = \sum_A (Z_A / |r - R_A|) - \sum_{ij} P_{ij} \int (\psi_i \psi_j / |r - r'|) dr'$$

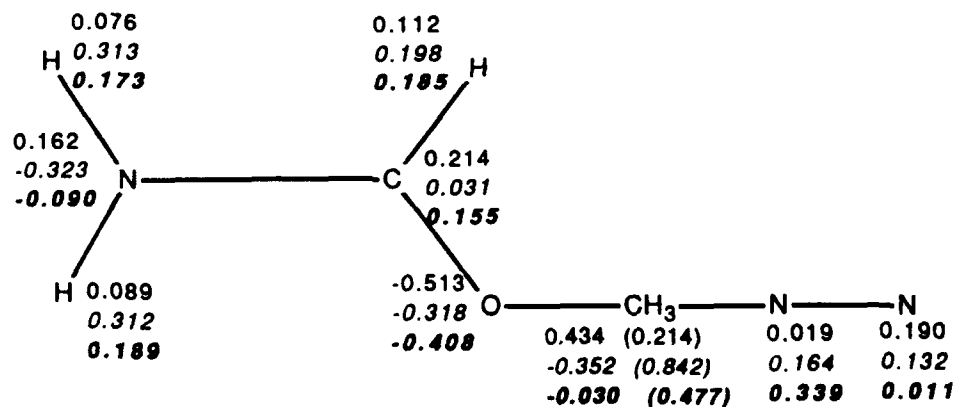
where Z_A is the charge on nucleus A, R_A is the position of nucleus A, and P_{ij} is the density matrix [18]. The method of Besler, Merz, and Kollman calculates $V(r)$ at a set of points on a molecular surface

Figure 5.4 - Calculated Atomic Charges in
Formamide-Based Systems



transition state for the
methylation of formamide
by the methyldiazonium ion (+1)

Legend:
PM3 (MOPAC 5.0)
DGauss
ESP (MOPAC 6.0)

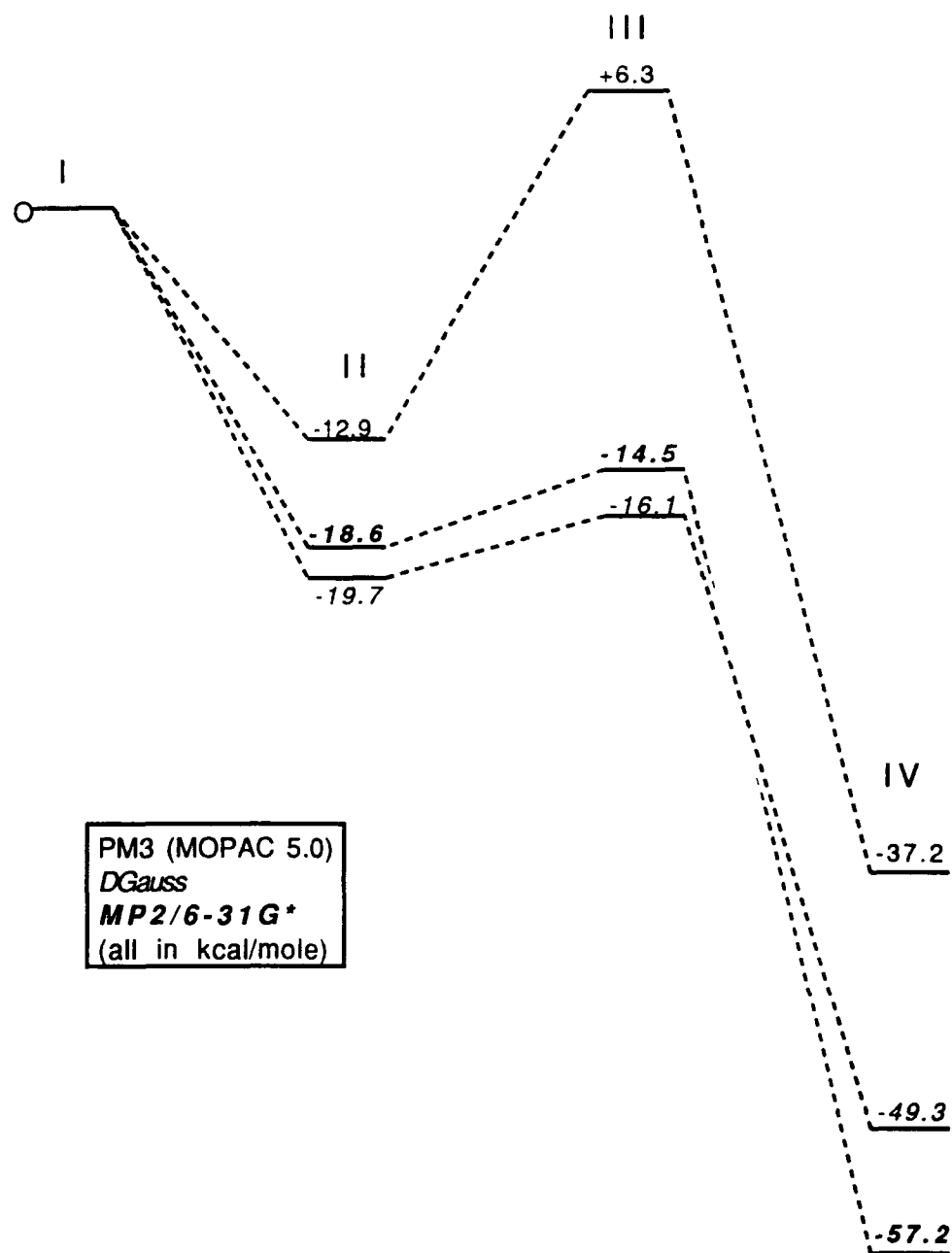


obtained as a combination of atomic surfaces determined by the atoms' van der Waals radii and densities after the method of Connolly [21]. Then, atomic charges, q_i , are assigned via a linear least squares fit procedure due to Chirlian and Francl [22] which minimizes $\sum_i^n (V_i - E_i)^2$ with $E_i = \sum_j^n (q_j / r_{ij})$ (where n is the number of charges) subject to the constraint that the set of q_i thus obtained satisfy $\sum_i^n q_i = q_{\text{total}}$. Such ESP-derived semiempirical charges have been found to compare favorably with similar ESP 6-31G* charges [18] which are, in turn, often more satisfactory than Mulliken charges [23].

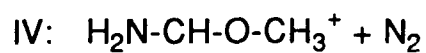
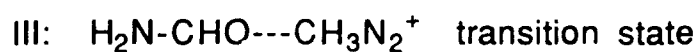
In Figure 5.4, there are significant differences in the three sets of calculated charges. From the discussion in the preceding paragraph, it seems that the ESP charges, since they are more rigorously defined, may be more reliable. However, in each method, the charge on the formamide oxygen is considerably less negative in the methylated form than in formamide alone. From this, it is reasonable that the difference in the interactions between the formamide oxygen and the water apparent in Figure 5.2 may be charge related. However, the reason for the lack of such an interaction at the transition state (Figure 5.3) is less obvious. Figure 5.4 shows that the formamide oxygen is most negative at the transition state. This phenomena will be addressed again in the discussion of analogous guanine-based calculations in the next section.

The MOPAC 5.0 and Gaussian 90 MP2/6-31G* values for the activation barrier (transition state relative to isolated reactants)

Figure 5.5 - Reaction Profile for the Methylation of Formamide by the Methyldiazonium Ion



PM3 (MOPAC 5.0)
DGauss
MP2/6-31G*
 (all in kcal/mole)



for the methylation of formamide by the methyldiazonium ion were reported in Chapter 3 to be 6.3 kcal/mole and -14.5 kcal/mole respectively. The corresponding DGauss result is -16.1 kcal/mole which is in reasonable agreement with the Gaussian 90 value. The reaction energy (enthalpy for MOPAC) profile calculated by MOPAC 5.0 PM3, DGauss, and Gaussian 90's MP2/6-31G* is shown in Figure 5.5. (The DGauss ion-molecule complex was optimized with the gradient convergence threshold set at "loose" instead of "medium". This changes the convergence criteria to 1×10^{-3} instead of 8×10^{-4} atomic units as required by the default "medium" option [6].) When a water molecule was included, as in Figure 5.3, the activation barriers (transition state relative to isolated reactants) were found to be 0.1 kcal/mole (MOPAC 5.0), -24.5 kcal/mole (DGauss), and -26.3 kcal/mole (MP2/6-31G*). Thus, the hydrogen bonding of the water to the H¹ hydrogen leads to a transition state stabilization of about 6 to 12 kcal/mole.

In the next section, calculations similar to those reported in this section are discussed but using guanine-based systems instead of formamide. The lesser computational scaling of DGauss allows *ab initio* level rigor to be applied to a guanine-sized system - a task impractical for Gaussian 90.

5.3 The Effect of Water on Guanine-Based Systems

The MDO6G (methyldiazonium ion attacking the O⁶ of guanine) reaction was chosen as the basis for studying the possible effect of water on the deprotonation of the N¹ nitrogen. The MOPAC 5.0 (Bartels's method) and MOPAC 6.0 (eigenvector following) results in

Chapter 4 for the MDO6G reaction were essentially identical. Force constant analysis in Chapter 4 indicated a well-characterized transition state. Also, the x-ray study of Parathasarathy and Fridley [2] provides an experimental basis for comparison for O⁶-methylguanosine.

MOPAC and DGauss calculations analogous to those of the previous section were performed for the MDO6G reaction. First, the geometries and charges predicted by MOPAC and DGauss for guanine, O⁶-methylguanine, and the MDO6G transition state were calculated. (In the DGauss MDO6G transition state optimization, the gradient convergence criteria was set to "loose".) Figures 5.6 and 5.7 show selected results of these calculations. All parameters reported in this section are the result of full geometry optimizations. Figure 5.7 includes ESP charges calculated with MOPAC 6.0 using PM3. Table 5.1 shows the complete set of heavy-atom bond lengths for guanine along with the MOPAC 5.0 PM3 and experimental values from Table 3.4 for comparison. In Table 5.2, a similar collection of bond lengths for the distal form of deprotonated (N¹) O⁶-methylguanine is compared with values from x-ray crystallography for deprotonated (N¹) distal O⁶-methylguanosine. The distal form was optimized for Table 5.2 in order to be consistent with the x-ray values. Elsewhere in this section, the proximal form is assumed.

The density-functional results of DGauss in Tables 5.1 and 5.2 are impressive. When compared with experiment, the bond length predicted by DGauss is superior to the PM3 prediction in every case. In about half the bond lengths, the DGauss prediction differs from the x-ray value by 0.01 Å or less. Thus, it seems that the

Figure 5.6 - Selected Optimized Geometry Parameters
for Guanine-Based Systems

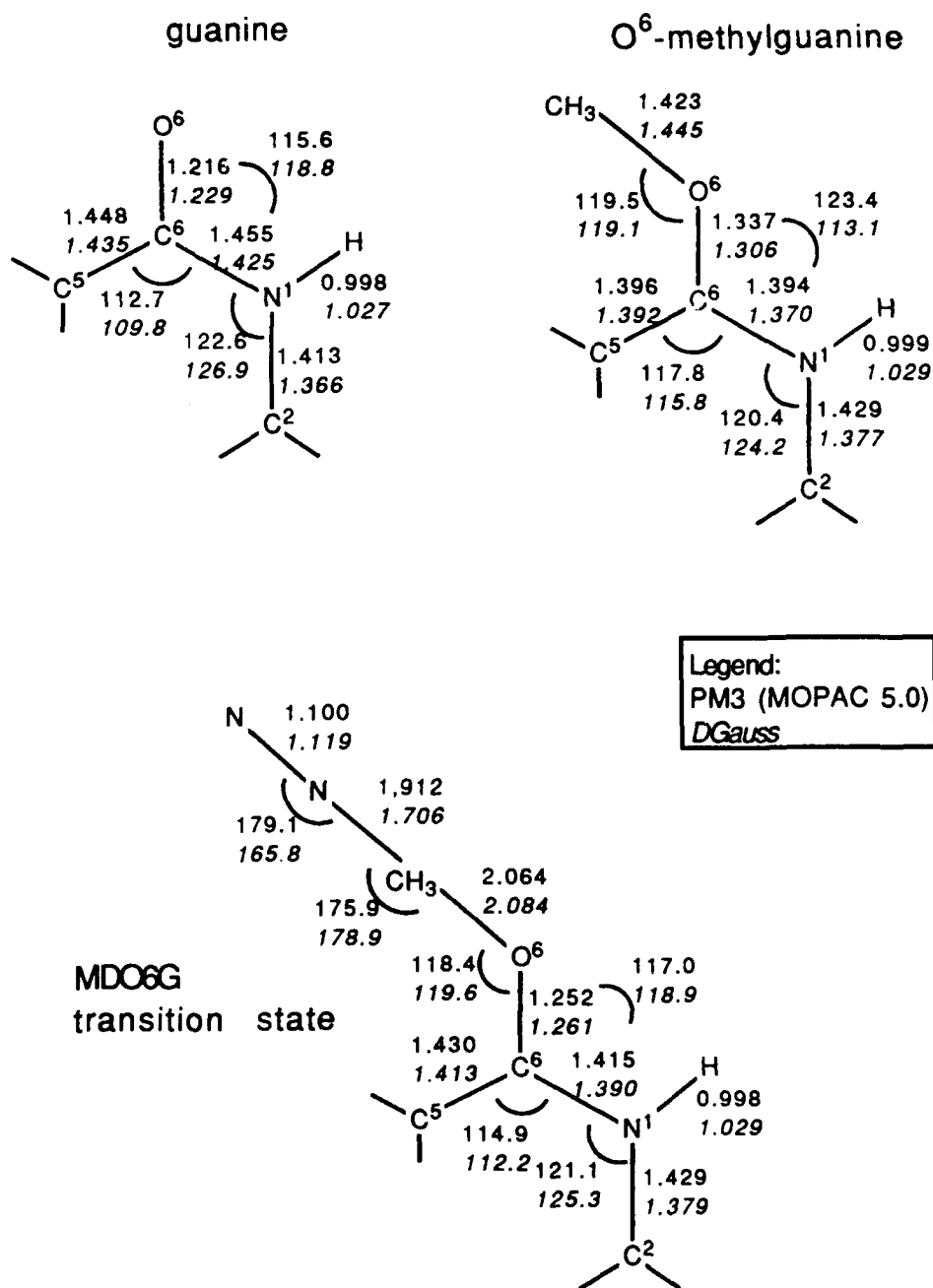


Figure 5.7 - Selected Charges in Guanine-Based Systems

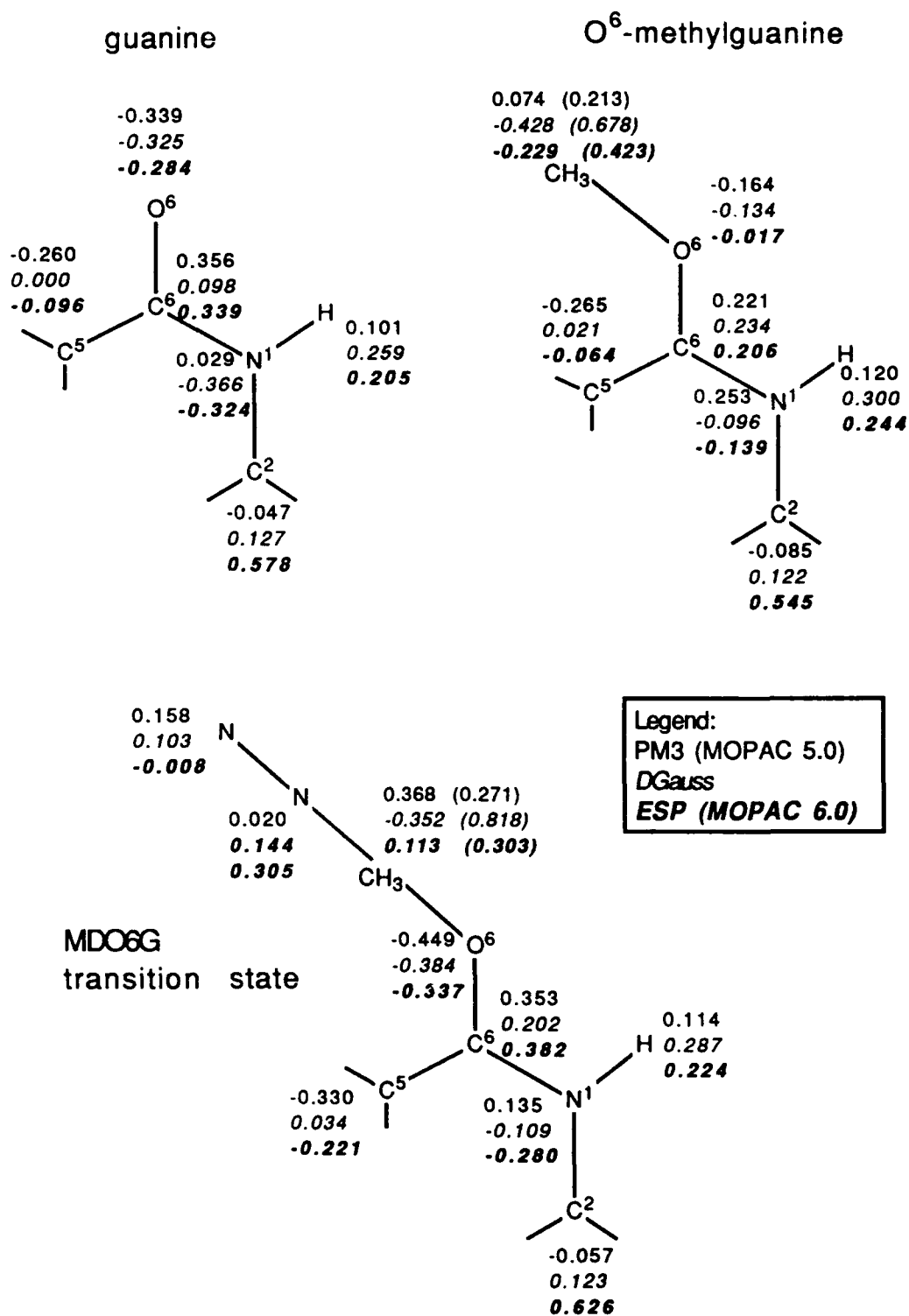


Table 5.1 - Bond Lengths in Guanine (Å)

<u>Bond</u>	<u>DGauss</u>	<u>PM3</u>	<u>Expt.</u> ¹
N ¹ -C ²	1.366	1.413	1.375
C ² -N ³	1.317	1.339	1.327
C ² -N ²	1.361	1.422	1.341
N ³ -C ⁴	1.350	1.398	1.355
C ⁴ -C ⁵	1.402	1.404	1.377
C ⁵ -C ⁶	1.435	1.448	1.415
C ⁶ -O ⁶	1.229	1.216	1.239
C ⁶ -N ¹	1.425	1.455	1.393
C ⁵ -N ⁷	1.373	1.397	1.389
N ⁷ -N ⁸	1.312	1.344	1.304
C ⁸ -C ⁹	1.380	1.411	1.374
C ⁴ -C ⁹	1.369	1.394	1.377

¹ From Reference 24, p. 52 (for 9-methylguanine).

Table 5.2 - Bond Lengths in Deprotonated (N¹)
Distal O⁶-Methylguanine (Å)

<u>Bond</u>	<u>DGauss</u>	<u>PM3</u>	<u>Expt.</u> ¹
N ¹ -C ²	1.353	1.396	1.362
C ² -N ³	1.341	1.363	1.342
C ² -N ²	1.362	1.404	1.354
N ³ -C ⁴	1.333	1.375	1.348
C ⁴ -C ⁵	1.409	1.418	1.384
C ⁵ -C ⁶	1.410	1.419	1.399
C ⁶ -O ⁶	1.335	1.353	1.338
C ⁶ -N ¹	1.325	1.352	1.311
C ⁵ -N ⁷	1.377	1.407	1.391
N ⁷ -N ⁸	1.313	1.336	1.301
C ⁸ -C ⁹	1.381	1.418	1.378
C ⁴ -C ⁹	1.372	1.399	1.371
O ⁶ -C(methyl)	1.423	1.414	1.447

¹ From Reference 2 for O⁶-methylguanosine.

geometrical parameters predicted in the structures in Figure 5.6 may be more reliable than the PM3 values. In view of the comparisons in Tables 5.1 and 5.2, this is certainly true of guanine and O⁶-methylguanine. There are, of course, no experimental structure parameters for the MDO6G transition state. However, it is noteworthy that the O⁶-C(methyl) and C(methyl)-N bond lengths predicted by DGauss for the MDO6G transition state are similar to the analogous bond lengths in the transition state for the methylation of formamide by the methyldiazonium ion (see Figure 5.1) as predicted by both DGauss and the MP2/6-31G* calculations.

The next step was to determine the effect of water on the guanine-based systems. PM3 geometry optimizations of guanine (MOPAC 5.0), O⁶-methylguanine (MOPAC 5.0), and the MDO6G transition state (MOPAC 6.0 - eigenvector following) each with a water molecule initially 1.5 Å from the N¹ hydrogen, and oriented as in the analogous calculations on formamide in the previous section, resulted in very little lengthening of the N¹-H bond. Similar results were obtained with DGauss (with the gradient convergence criteria set to "loose" for the transition state optimization as described earlier). Selected optimized parameters are shown in Figures 5.8 and 5.9. Figure 5.8 shows significant differences in the optimized position and orientation of the water molecule as part of the guanine structure *versus* the methylated guanine structure. In the MOPAC guanine structure, the water is angled towards the O⁶ oxygen by about 20° relative to its starting position of being colinear with the N¹-H bond. In the MOPAC results, the water hydrogens end up out of the plane of the ring system - one by 20.5° and the other by 94.8°.

Figure 5.8 - The Effect of Water on Guanine
and O⁶-Methylguanine (+1)

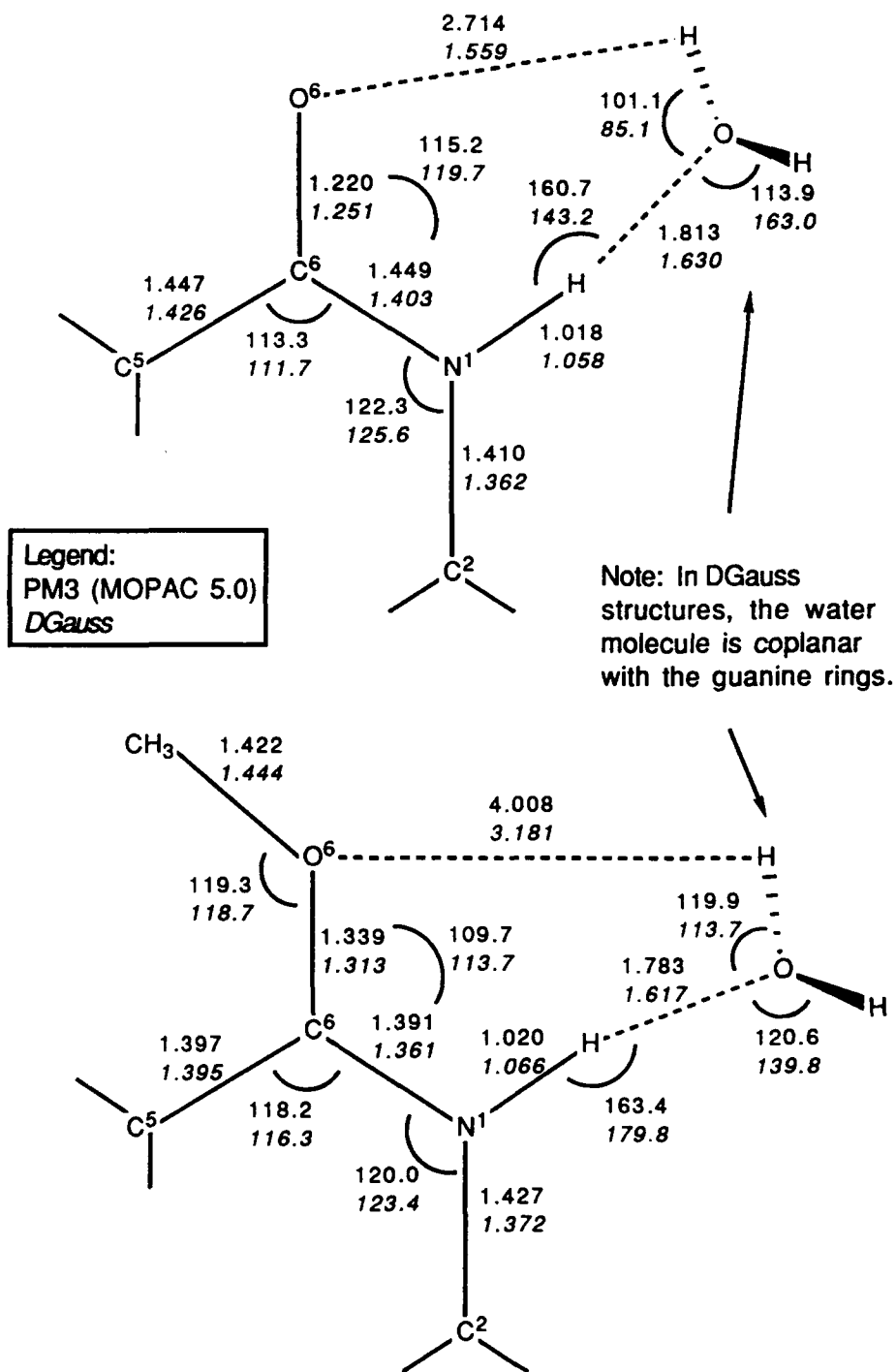
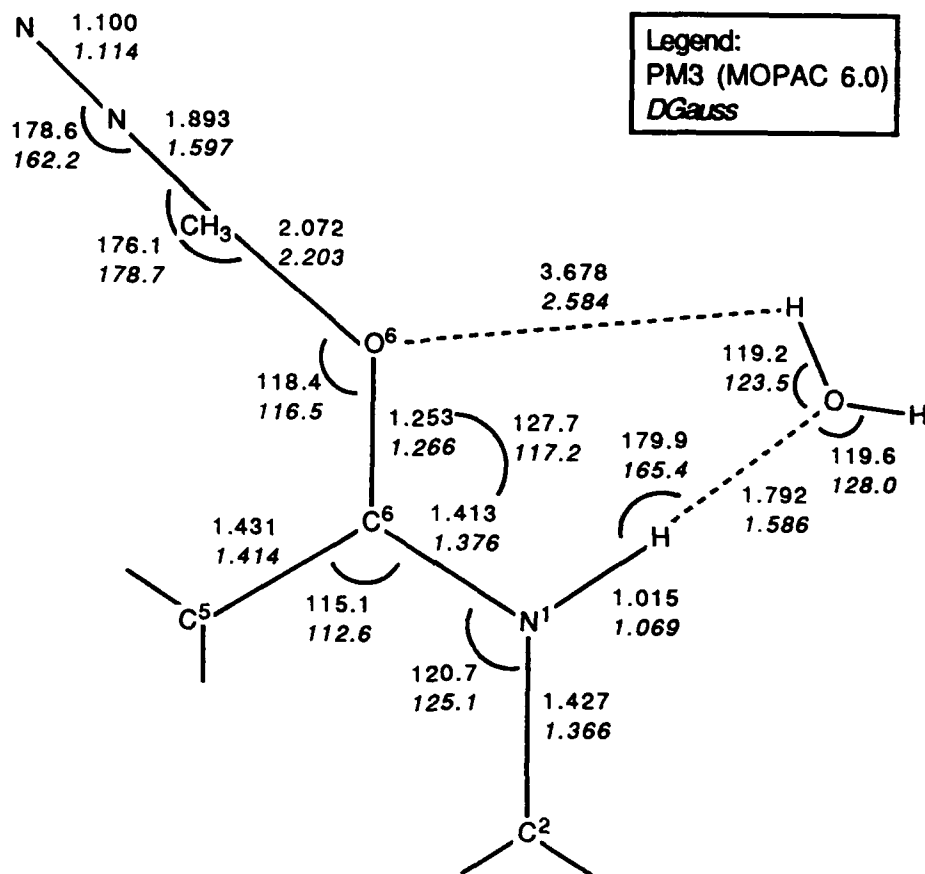
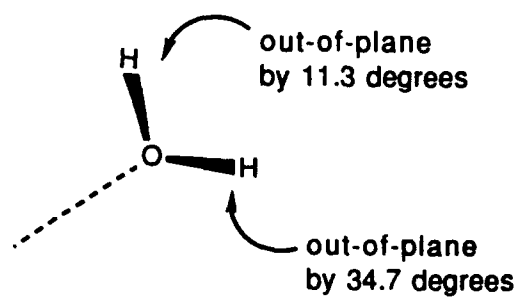


Figure 5.9 - The Effect of Water on the MDO6G Transition State (+1) - Results of Transition State Optimization



Note: In DGauss structure, the water hydrogens are out of the plane as:



The one closer to the plane is the one closer to the O⁶ oxygen at an O⁶...H distance of 2.714 Å. In contrast, the optimized position of the water as part of the methylated guanine structure has the water bent away from the O⁶ oxygen by nearly 20°. One water hydrogen is out of the plane by 39.3° and the other by 99.6°. The former is the closer one to the O⁶ oxygen at 4.008 Å. The DGauss results differ similarly with an O⁶...H distance of 1.559 Å in the guanine structure and 3.181 Å in the methylated structure, but the water hydrogens remain coplanar with the rings and the water oxygen remains colinear with the N¹-H bond in the methylated structure.

As in the formamide model studies in the previous section, the reason for these differences is apparently the change in the charge distribution. The ESP (MOPAC 6.0) calculated charges for both systems as well as for the MDO6G transition state (with hydrogens summed into heavy atoms) are shown in Table 5.3. In the methylated guanine, the O⁶ oxygen is significantly less negative and, therefore, less likely to interact with the water. At the same time, the N¹-H region becomes more positive upon methylation causing a stronger interaction with the water oxygen. Figures 5.10 and 5.11 show the DGauss-calculated electron densities (contour lines spaced at intervals of about 0.5 electrons/Å³), in the plane of the rings, for guanine and O⁶-methylguanine. Electron density comparisons using these figures, produced by the UniChem package [6], should be of relative electron density in one structure versus relative electron density in another structure. For example, one could compare the N¹ relative to C⁵ in Figure 5.10 *versus* the N¹ relative to C⁵ in Figure 5.11. Since the maximum electron density (i.e., the darkest shading)

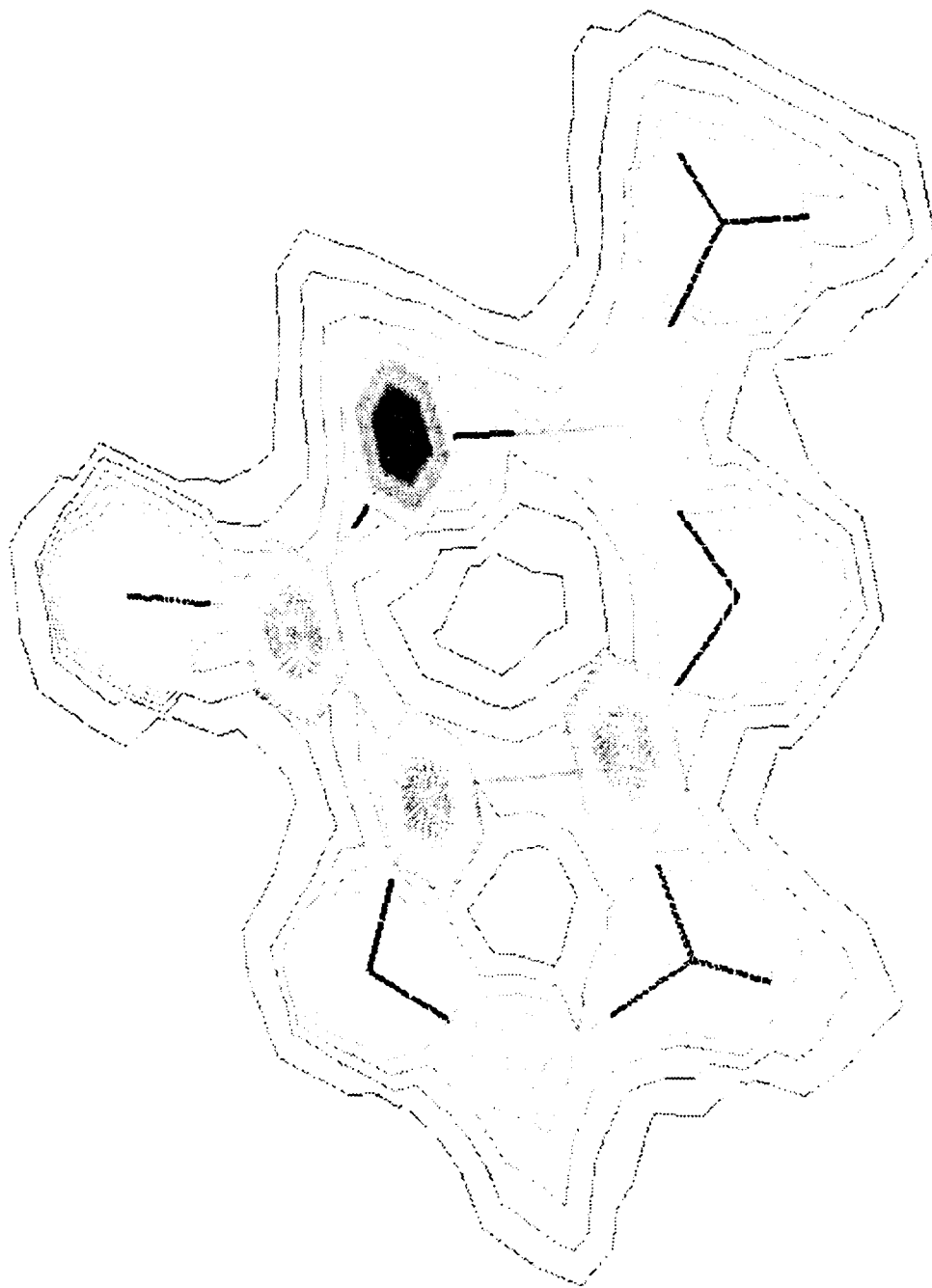
is different in the two figures, due largely to the charge difference, direct comparisons of, for example, C⁵ in Figure 5.10 with C⁵ in Figure 5.11, are not useful. The methylation-induced decrease in electron density (relative to the rest of the molecule) is apparent in the C⁶-N¹-H region.

Table 5.3 - ESP (MOPAC 6.0) Calculated Charges in Guanine, O⁶-Methylguanine (+1), and the MDO6G Transition State (+1)

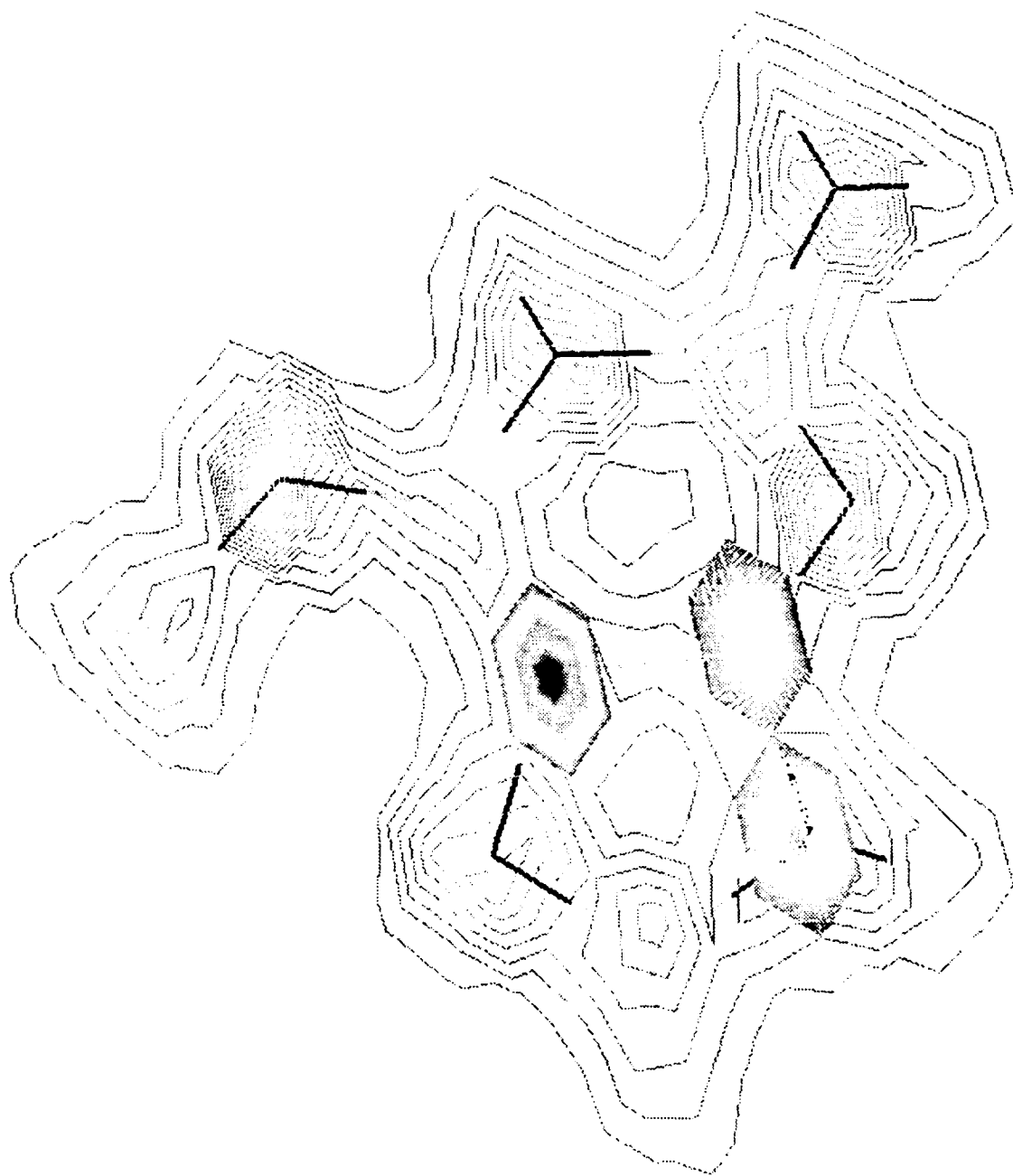
<u>Atom</u>	<u>Guanine</u>	<u>O⁶-Methylguanine</u>	<u>MDO6G TS</u>
N ¹ (+H)	-0.119	0.105	-0.056
C ²	0.578	0.545	0.626
N ² (+2H)	-0.034	0.125	0.054
N ³	-0.623	-0.566	-0.613
C ⁴	0.284	0.324	0.289
C ⁵	-0.096	-0.064	-0.221
C ⁶	0.339	0.206	0.382
O ⁶	-0.284	-0.017	-0.337
N ⁷	-0.348	-0.320	-0.206
C ⁸ (+H)	0.115	0.199	0.038
N ⁹ (+H)	0.188	0.271	0.332
C _m (+3H)	-	0.194	0.416
2N (diazonium)	-	-	0.297

The charges in Table 5.3 show that the only two positions which experienced a decrease in charge upon methylation were C² and C⁶. The atomic charge of each of the other positions increased as the +1 charge of the methyldiazonium ion became part of the system. The two largest increases were at O⁶ and N¹ which is consistent with the reduced interaction of the water with O⁶ upon methylation. The increase in charge at N² could account for the

Figure 5.10 - DGauss Electron Density Diagram for Guanine



**Figure 5.11 - DGauss Electron Density Diagram
for O⁶-Methylguanine (+1)**

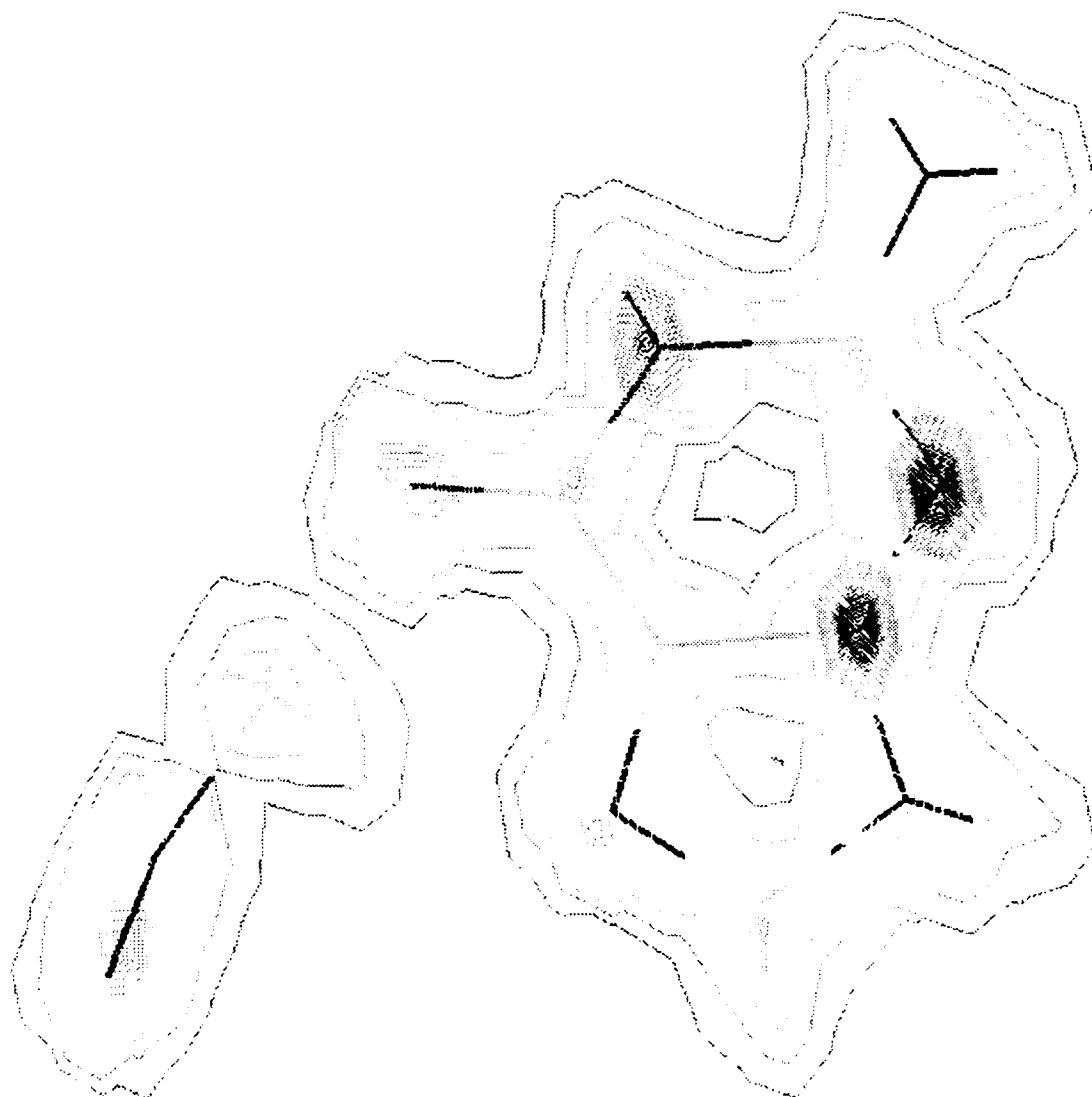


water being bent away from the O⁶ oxygen in the methylated form in Figure 5.8.

As in the analogous formamide calculations, it is unclear from the atomic charges in the MDO6G transition state why the water interacts relatively little, if any, with the O⁶ oxygen in Figure 5.9. From Table 5.3, the charge on the O⁶ oxygen is most negative at the transition state. The DGauss results in Figure 5.9 show some O⁶...water interaction, but not as much as might be expected from the charge on O⁶. Some explanation may be drawn from the electron density diagram for the MDO6G transition state shown in Figure 5.12 (with contour lines again at intervals of about 0.5 electron /Å³). The closeness of the electron density contours around O⁶ do not change very much as the reaction proceeds as shown in Figures 5.10, 5.12, and 5.11. However, relative to O⁶, there are considerable changes at N¹. At the transition state, the relative electron density near N¹ has decreased appreciably making it more attractive to the water oxygen. The apparent reduced interaction of the water with O⁶ at the transition state may be more due to increased attractiveness of N¹ and other parts of the system, like N², instead of decreased attractiveness of O⁶.

The relatively significant increase in atomic charge in guanine's N¹-H region upon methylation should facilitate loss of the N¹ hydrogen to produce the deprotonated form observed by Parathasarathy and Fridey [2] and effect the eventual formation of a hydrogen-bonded O⁶-methylG:T base pair as described by Leonard, *et. al.* However, not one of the calculations so far in this chapter or in the previous chapter indicate any significant lengthening of the

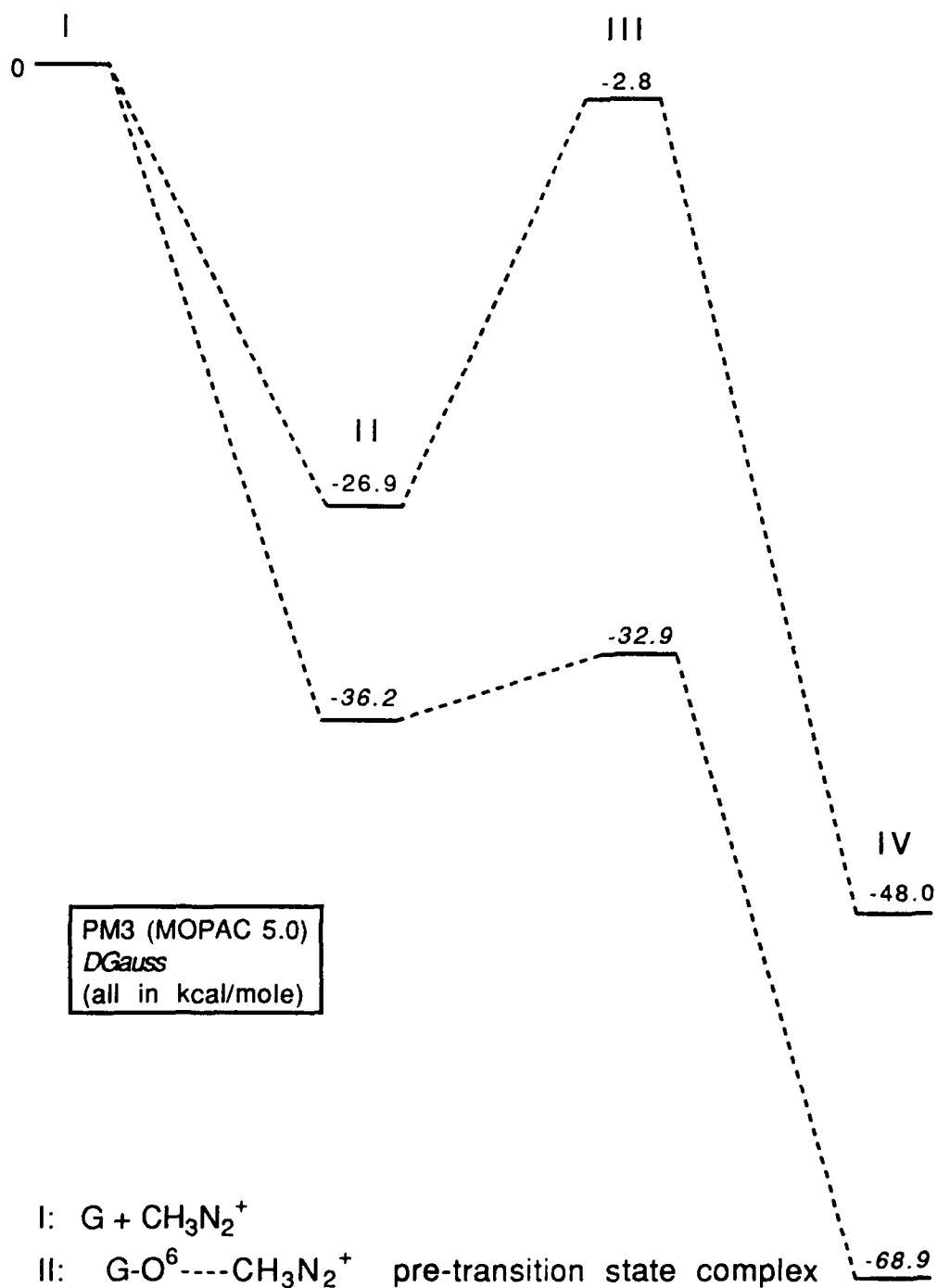
**Figure 5.12 - DGauss Electron Density Diagram for
the MDO6G Transition State (+1)**



N¹-H bond upon alkylation either with or without the influence of a water molecule. Since the deprotonation did not occur spontaneously in these calculations, it was next necessary to force the loss of the hydrogen and study the energetics involved. In the next section, the results of MOPAC PM3 calculations on the forced abstraction of the N¹ hydrogen are reported.

From the DGauss energies of the reactants, products, and the transition state, the DGauss results for the ΔE and activation energy of the MDO6G reaction are -68.9 kcal/mole and -39.9 kcal/mole respectively. These values are significantly lower than the corresponding PM3 results in Tables 4.7 and 4.13. Since DGauss results in this study have been generally superior to PM3 results when compared to Gaussian 90 MP2/6-31G* calculations (see previous section), the DGauss values for the *guanine* reaction may have more credibility. The reaction profile is shown in Figure 5.13 as calculated by MOPAC 5.0 PM3 and DGauss. (The DGauss ion-molecule complex and transition state were optimized with the gradient convergence threshold set to "loose".) The MOPAC results are consistently above the DGauss results as they were, also, in the formamide reaction (see Figure 5.5). This consistency, admittedly from a very small sampling, indicates, however, that the comparisons of PM3 activation parameters in Chapter 4 for alkylation by alkyldiazonium ions and carbocations are not invalid. Even if the PM3 parameters are not correct individually, comparisons among the various results are still useful and, indeed, were reasonably consistent with experimental observations as described in the "Conclusions" section of Chapter 4.

Figure 5.13 - Reaction Profile for the O⁶ Methylation of Guanine by the Methyldiazonium Ion



I: $G + CH_3N_2^+$

II: $G-O^6 \cdots CH_3N_2^+$ pre-transition state complex

III: $G-O^6 \cdots CH_3N_2^+$ transition state

IV: O^6 -methylG⁺(with N¹ hydrogen) + N₂

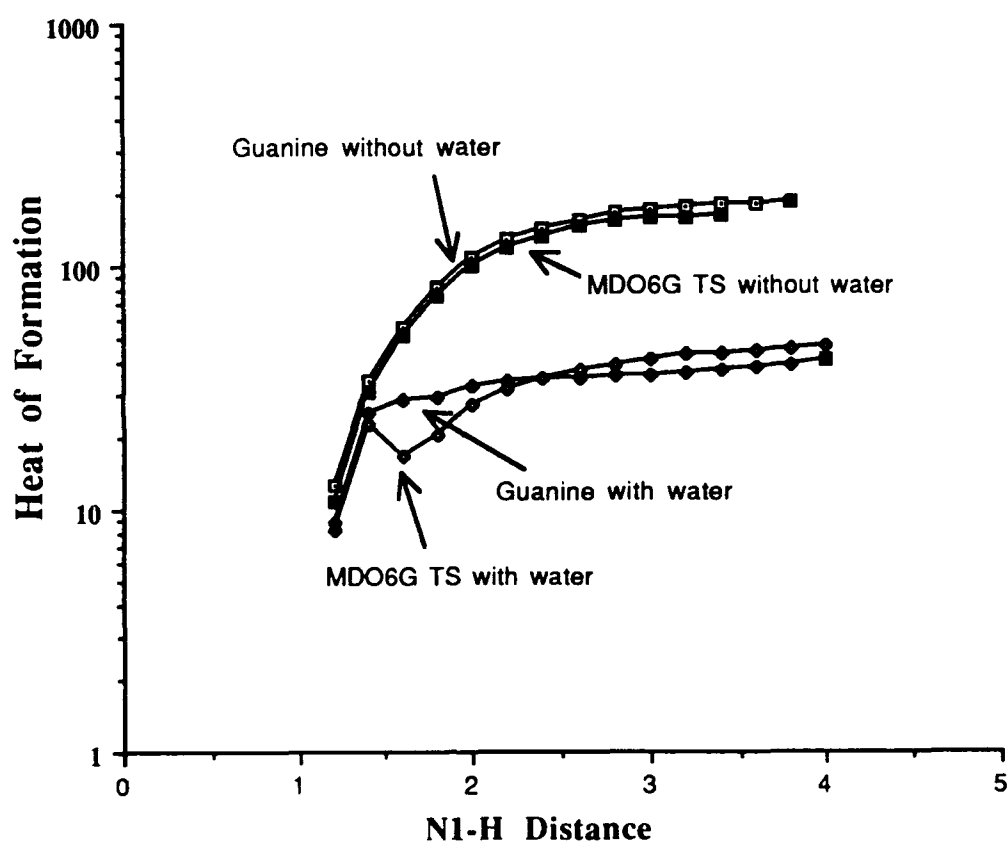
5.4 - Forced Deprotonation of the N¹ Nitrogen

In this section, comparisons are made in the relative ease of proton loss at N¹ among O⁶-methylated and nonmethylated guanine with and without the presence of water. The procedure was to perform MOPAC 5.0 PM3 coordinate-driving calculations in which the N¹-H bond is incrementally stretched with full energy minimization of all other geometrical parameters at each fixed N¹-H distance. This calculation was performed for each of four cases - guanine with water, guanine without water, the MDO6G transition state with water, and the MDO6G transition state without water. In the calculations which included water, the water molecule was initially 1.5 Å from the N¹ hydrogen and oriented as described in the previous section.

The calculations began at the equilibrium bond length of 0.998 Å and ended at a maximum of 4.0 Å. The heat of formation *versus* N¹-H distance provided the data for approximate reaction coordinate diagrams for proton abstraction. The resulting four sets of data are shown plotted in Figure 5.14. In the first increment of proton abstraction from the MDO6G transition state (with or without water) the N₂ molecule was found to be about four angstroms from the ring system and did not appear to interact any further during the stretching of the N¹-H bond. Thus, the MDO6G transition state calculations are essentially equivalent to proton abstraction from the methylated product. To facilitate comparisons among the four curves, each set of data was shifted to a common zero point.

It is immediately obvious from Figure 5.14 that in only one case, the MDO6G transition state with water, was there an enthalpy

Figure 5.14 - Heat of Formation *versus* N¹-H Distance for Forced Deprotonation at N¹



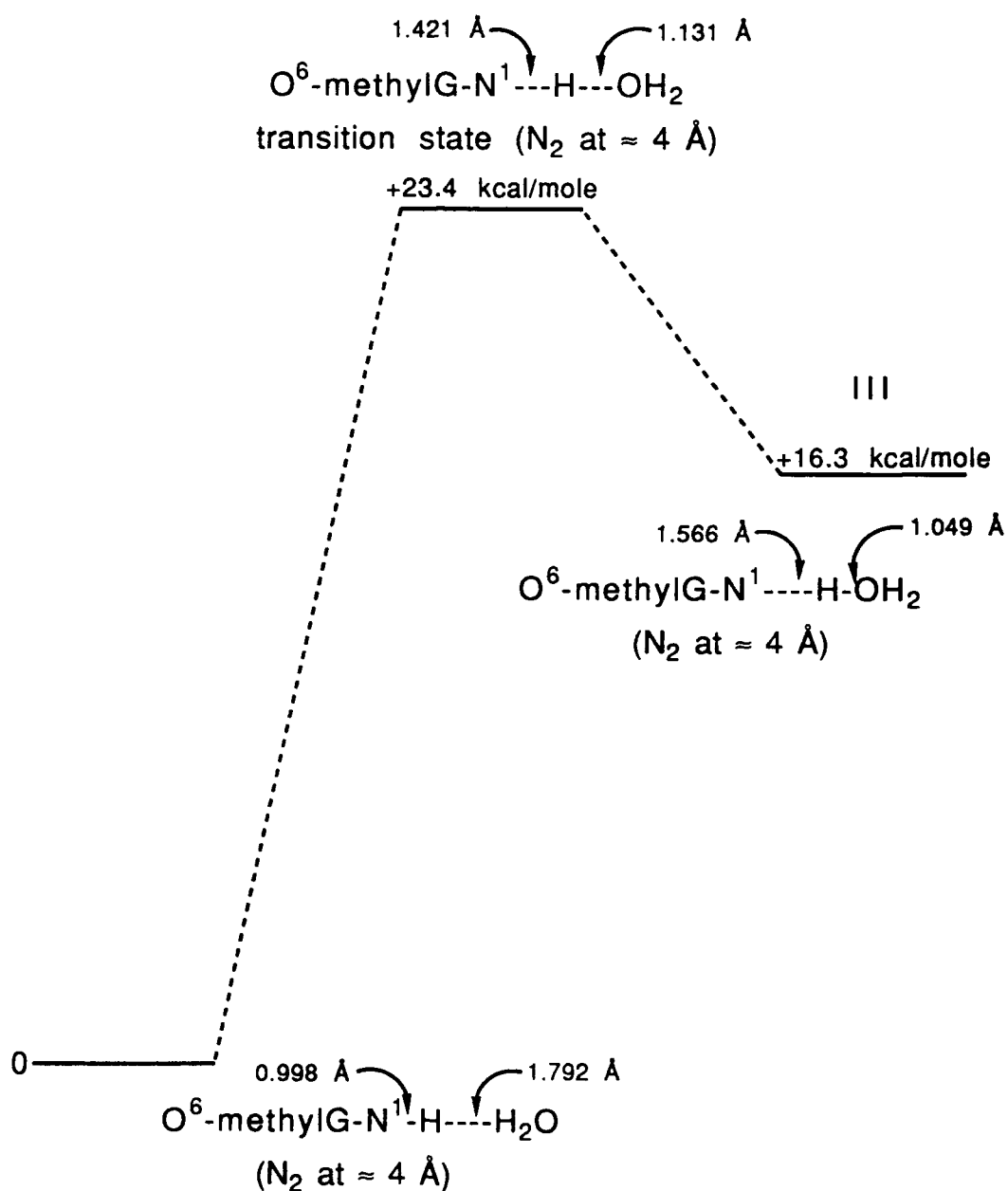
minimum reached at an N¹-H distance greater than the equilibrium bond length. This indicates that loss of the proton at N¹ is not likely to occur unless two factors are included - methylation of guanine's O⁶ oxygen and the presence of a proton acceptor like water. Thus, it seems reasonable that, since imminent proton loss was not apparent in either the transition state or the methylated product, deprotonation may occur in another step separate from the original alkylation and that the deprotonation involves a proton acceptor, very possibly water. How this fits into the overall mutagenic mechanism is described in the next section.

The structures corresponding to the enthalpy well and the apparent maximum in the "MDO6G TS with water" curve in Figure 5.14 were minimized using MOPAC 5.0 PM3. The resulting reaction profile is shown in Figure 5.15. The activation enthalpy is +23.4 kcal/mole and the enthalpy change for the reaction which ends with the H₃O⁺ hydrogen-bonded to N¹ is +16.3 kcal/mole.

5.5 - Conclusions

From the results in Chapter 4, it was apparent that the O⁶-alkylation of guanine is not sufficient to cause loss of the proton at N¹. None of the products or transition states for the alkylation of guanine at O⁶ showed any tendency to lengthen the N¹-H bond to the point where the proton would be lost. However, as noted in Chapter 4, one reason for this result could have been that the semiempirical treatment was inadequate in describing the making and breaking of bonds in these reactions. Similar results had been obtained with MOPAC and Gaussian 90 MP2/6-31G* calculations on the model

Figure 5.15 - MOPAC Reaction Profile for the Loss
of the N¹ Hydrogen to a Water Molecule



formamide reaction. Unfortunately, the application of traditional *ab initio* techniques to a guanine-sized system is currently impractical.

For this reason, the use of density-functional theory (DFT) was explored in this chapter. Used for years in solid state physics and materials science, DFT is only recently becoming more widely used by chemists. Some reports from DFT calculations indicate that DFT predictions are comparable with Hartree-Fock results which include some intermediate level of electron correlation treatment [25]. Indeed, in this chapter, the DGauss density-functional results were in line with the MP2/6-31G* results for the formamide systems considered. It was the intent of this approach to determine whether or not the electron correlation treatment inherent in density-functional theory would allow N¹ deprotonation of guanine upon O⁶ alkylation.

It is instructive overall, and reassuring for the PM3 results, that the density-functional results agreed qualitatively with the results in Chapter 4. The DGauss calculations showed no tendency for N¹ deprotonation during the O⁶ methylation of guanine by the methyldiazonium ion. There was little or no lengthening of the N¹-H bond either at the transition state or in the methylated product. Thus, it seemed that something else was necessary for the proton to be lost. Since the inclusion of water in the calculation of the proton loss in Chapter 4 reduced the endothermicity of the process, the effect of water on the guanine reaction and the model formamide reaction was addressed with MOPAC and DGauss.

The results of the calculations involving water in sections 5.2

and 5.3 showed that a water molecule in the vicinity of the N¹-H region interacted significantly with the O⁶ oxygen until the methylation took place. Then, the water molecule tended to become more closely associated with the N¹-H region. However, even during (i.e., at the transition state) and after the methylation, the attraction of the water molecule for the now more positive N¹ hydrogen was not sufficient to effect deprotonation. When the forced abstraction of the proton was studied in section 5.4, it became apparent that two factors contributed to the ease of proton loss - the O⁶ methylation and the influence of the water molecule. Thus, it may be that the mechanism which leads to an N¹ deprotonated O⁶-alkylguanine as part of a DNA sequence, which could then errantly hydrogen-bond with a thymine, may occur as a two-step process - first, the alkylation and second, the proton loss to a proton acceptor like water as depicted in Figures 5.13 and 5.15.

Water molecules would normally be scarce inside the double helix due to the hydrophobic interactions of the rings being stacked on top of each other and the competition for hydrogen-bonding sites of complementary base pairs *versus* water [26]. However, during replication or transcription, when the double strand is "unwound", water molecules would have more potential access to hydrogen bond at sites normally inside the helix such as the N¹ hydrogen. It is easily imagined that, as the two single strands recombine into the helix, the competition for hydrogen bonding to the N¹ hydrogen could result in loss of the proton if that loss were made energetically easier because of alkylation at the O⁶ oxygen and the reduced interaction of the water with the O⁶ oxygen due to the alkylation.

Once the deviant base was incorporated into the double helix, it could code for thymine in the next replication. However, this last sequence of events is speculation. Nowhere in the calculations in this study have the effects of neighboring bases or the rest of a double helix been considered. Indeed, the treatment of the water influence was somewhat artificial. However, the results of these calculations strongly indicate a two-step process.

There are also lessons of a computational nature. It should be pointed out that some of the results reported in this chapter are from "incomplete" optimizations. As in section 4.3, the MOPAC 5.0 transition state optimizations involving guanine used Bartels's method and were judged complete not when the MOPAC criterion had been met, but when the heats of formation and geometries were changing negligibly in consecutive job runs. The DGauss optimizations of the MDO6G transition state with a water molecule in Figure 5.9 and of the ion-molecule complexes of Figures 5.5 and 5.13 were not converged after considerable CPU time. Examination of the output files showed that the largest component of the gradient remained in the 10^{-2} to 10^{-3} (atomic units) range. Since it appeared that even the "loose" criterion would not be met and since the geometries were fairly stable, the optimizations were judged complete. Nevertheless, the success of the DGauss density-functional calculations on the formamide models and, therefore, the confidence with which it could be applied to the guanine reaction, is worth noting. With density-functional theory, chemists have a chance to apply the rigor of traditional *ab initio* calculations to systems too large to currently be practically

considered for traditional *ab initio* techniques. However, as with any computational method, density-functional theory must be judiciously used. It will have its strengths and weaknesses which will be identified as more and more comparisons are made among density-functional results, other computational techniques, and, the really critical comparison, with experimental observations. Hopefully, this study will be a part of that learning process.

In addition, the success of the semiempirical techniques of MOPAC in characterizing the transition states and other species in this study will add to the body of knowledge on semiempirical techniques, especially the new PM3 method. Although, DGauss was successfully used for systems too large for traditional *ab initio* techniques, it still was computationally more expensive than MOPAC. It should not be surprising that the DGauss transition state results compared more favorably with the Gaussian 90 results than did those of MOPAC since MOPAC is parameterized for stable molecules instead of transition states while DGauss is an *ab initio* -type technique. Nevertheless, the MOPAC results in this chapter were qualitatively comparable to those of DGauss and required considerably less time. Semiempirical techniques provide a relatively fast and inexpensive way to achieve meaningful results both as answers in themselves and as a guide to the use of other methods like density-functional theory. There will continue to be a need for a balance of various computational methods, each with its own strengths and weaknesses, to complement each other and experimental work.

References

1. Loveless, A. (1969) *Nature*, 223, 206-207.
2. Parathasarathy, R. and Fridey, S.M. (1986) *Carcinogenesis*, 7, 221-227.
3. Leonard, G.A., Thomson, J., Watson, W.P., and Brown, T. (1990) *Proceedings of the National Academy of Sciences*, 87, 9573-9576.
4. Stewart, J.J.P. (1989) MOPAC Manual (Fifth Edition) A General Molecular Orbital Package, Air Force Systems Command, United States Air Force.
5. Stewart, J.J.P. (1990) MOPAC Manual (Sixth Edition) A General Molecular Orbital Package, Air Force Systems Command, United States Air Force.
6. *UniChem Chemistry Codes - SG-5151 1.0* (1990) Cray Research, Inc.
7. Becke, A.D. (1988) *Physical Review*, 38, A3098.
8. Becke, A.D. (1988) *Journal of Chemical Physics*, 88, 2547.
9. Becke, A.D. (1989) in The Challenge of d and f Electrons: Theory and Computation, Salahub, D.R. and Zerner, M.C., eds., ACS Symposium Series 394, p. 166.
10. Perdew, J.P. (1986) *Physical Review*, 33, B8822.
11. Andzelm, J. (1991) in Density Functional Methods in Chemistry, Labanowski, J.K. and Andzelm, J.W., eds., Springer-Verlag, New York.
12. Schlegel, H. B. (1987) in Ab Initio Methods in Quantum Chemistry-I, Lawley, K.P., ed., Wiley & Sons, pp. 249-285.
13. McIver, J.W. and Komornicki, A. (1971) *Chemical Physics Letters*, 10, 303.

14. Ford, G.P. and Scribner, J.D. (1983) *Journal of the American Chemical Society*, 105, 349-354.
15. *Gaussian 90*, M.J. Frisch, M. Head-Gordon, G.W. Trucks, J.B. Foresman, H.B. Schlegel, K. Raghavachari, M.A. Robb, J.S. Binkley, C. Gonzalez, D.J. Defrees, D.J. Fox, R.A. Whiteside, R. Seeger, C.F. Melius, J. Baker, R.L. Martin, L.R. Kahn, J.J.P. Stewart, S. Topiol, and J.A. Pople, Gaussian, Inc., Pittsburgh, PA 1990.
16. Bartels, R.H. (1972) *University of Texas, Center for Numerical Analysis, Report CNA-44*, Austin, TX.
17. Baker, J. (1986) *Journal of Computational Chemistry*, 7, 385-395.
18. Besler, B.H., Merz, K.M., and Kollman, P.A., *Journal of Computational Chemistry*, 11, 431-439.
19. Hehre, W.J., Radom, L., Schleyer, P.v.R., and Pople, J.A., *Ab Initio Molecular Orbital Theory*, Wiley & Sons, New York.
20. Scrocco, E. (1978) *Advances in Quantum Chemistry*, 11, 115.
21. Connolly, M.L. (1983) *Journal of Applied Crystallography*, 16, 548.
22. Chirlian, L.E. and Francl, M.M. (1987) *Journal of Computational Chemistry*, 8, 894.
23. Cox, S.R. and Williams, D.E. (1981) *Journal of Computational Chemistry*, 2, 304.
24. Saenger, W. (1984) *Principles of Nucleic Acid Structure*, Springer-Verlag, New York.
25. Borman, S. (1990) *Chemical & Engineering News*, April 9, 22-30.
26. Lewin, B. (1990) *Genes IV*, Oxford University Press, Oxford.

APPENDIX

In this appendix are shown several representative types of input files pertinent to this study. Each file is provided with a heading to describe its purpose.

MOPAC data files begin with a set of keywords which control the type of calculation being performed and the output desired. Following the keywords are a descriptive line, a blank line, and the input geometry in internal coordinates (cartesian coordinates are also accepted) [1]. The coordinate-driving calculation provides an approximate heat of formation versus reaction coordinate profile. As described in Chapter 2, a SADDLE calculation is then performed to locate an approximate transition state. The resulting structure is then refined in a subsequent NLLSQ (Bartels's method - MOPAC 5.0 [1]) or TS (eigenvector following - MOPAC 6.0 [2]) calculation. Also shown is a MOPAC 6.0 data file for a simple EF (eigenvector following) geometry optimization.

Two Gaussian 90 [3] input files are shown. The first is an MP2/6-31G* energy minimization of formamide. The second is an HF/6-31G* optimization of the transition state for the methylation of formamide by the methyldiazonium ion. Active geometry variables are in the last section of each file.

The two DGauss input files shown, one for energy minimization and one transition state optimization, were generated with UniChem 1.0 [4]. The UniChem interfaces through a Silicon Graphics IRIS workstation to provide menu-assisted molecular building and setup/control of input files.

Coordinate-driving calculation for the methyldiazonium ion
attacking the O⁶ position of guanine (MDO6G)

PM3 PRECISE T=1000M MMOK NOXYZ NOINTER CHARGE=+1
PM3 - METHYLDIAZONIUM ION ATTACKING O6 OF GUANINE

N	0.000000	0	0.000000	0	0.000000	0	0	0	0
C	1.394077	1	0.000000	0	0.000000	0	1	0	0
C	1.404325	1	106.801378	1	0.000000	0	2	1	0
N	1.396862	1	108.453460	1	0.108373	1	3	2	1
C	1.343529	1	108.021841	1	-0.048862	1	4	3	2
C	1.448177	1	120.264924	1	179.666046	1	3	2	1
O	1.215720	1	131.742377	1	-0.058815	1	6	3	4
N	1.454851	1	112.731107	1	0.824810	1	6	3	2
C	1.412778	1	122.613229	1	-2.093143	1	8	6	3
N	1.339289	1	123.867170	1	2.988359	1	9	8	6
N	1.421628	1	118.551191	1	-172.323598	1	9	8	6
H	0.995888	1	113.378003	1	-41.740171	1	11	9	8
H	0.995845	1	113.624355	1	-171.084805	1	11	9	8
H	0.998327	1	118.455444	1	2.387119	1	8	9	11
H	1.093162	1	125.310384	1	-179.952290	1	5	4	3
H	0.987213	1	126.422138	1	0.379054	1	1	2	10
C	6.000000	-1	180.000000	1	0.000000	1	7	6	3
H	1.000000	1	70.000000	1	0.000000	1	17	7	3
H	1.000000	1	70.000000	1	120.000000	1	17	6	3
H	1.000000	1	70.000000	1	-120.000000	1	17	6	3
N	1.400000	1	180.000000	1	0.000000	1	17	7	3
N	1.100000	1	180.000000	1	0.000000	1	21	17	3
O	0	0	0	0	0	0	0	0	0

4 3 2.5 2.3 2.1 1.9 1.7 1.5 1.4 1.3 1.2

**SADDLE calculation for the MDO6G reaction (locates
approximate transition state)**

PM3 PRECISE T=1000M MMOK NOXYZ NOINTER SADDLE XYZ CHARGE=+1
SADDLE CALCULATION FOR METHYLDIAZONIUM ION ATTACKING O6 OF GUANINE

N	0.000000	0	0.000000	0	0.000000	0	0	0	0
C	1.397273	1	0.000000	0	0.000000	0	1	0	0
C	1.402995	1	106.454492	1	0.000000	0	2	1	0
N	1.404436	1	108.816434	1	-0.030385	1	3	2	1
C	1.347830	1	107.530487	1	0.012150	1	4	3	2
C	1.440456	1	120.762833	1	-179.515631	1	3	2	1
O	1.233305	1	128.135017	1	0.034693	1	6	3	4
N	1.429102	1	113.879472	1	0.075258	1	6	3	2
C	1.423805	1	121.652279	1	-1.226548	1	8	6	3
N	1.348205	1	123.791595	1	2.364658	1	9	8	6
N	1.402259	1	118.977041	1	-172.163050	1	9	8	6
H	0.994046	1	115.613729	1	-33.603524	1	11	9	8
H	0.994725	1	115.421316	1	-170.202041	1	11	9	8
H	0.998457	1	119.137968	1	5.406088	1	8	9	11
H	1.095799	1	125.165631	1	-179.882382	1	5	4	3
H	0.988885	1	126.522887	1	0.510200	1	1	2	10
C	2.500000	1	113.751249	1	-3.176614	1	7	6	3
H	1.161639	1	82.326975	1	1.130931	1	17	7	3
H	1.121010	1	53.296442	1	143.421705	1	17	6	3
H	1.102200	1	94.747900	1	-107.167995	1	17	6	3
N	1.413226	1	149.172953	1	121.146385	1	17	7	3
N	1.107199	1	178.532201	1	15.203747	1	21	17	3
O	0.000000	0	0.000000	0	0.000000	0	0	0	0
N	0.000000	0	0.000000	0	0.000000	0	0	0	0
C	1.387669	1	0.000000	0	0.000000	0	1	0	0
C	1.430250	1	106.710008	1	0.000000	0	2	1	0
N	1.413385	1	107.667043	1	0.028407	1	3	2	1
C	1.331647	1	108.206435	1	0.012705	1	4	3	2
C	1.399378	1	119.141719	1	-179.882219	1	3	2	1
O	1.324337	1	134.249678	1	-0.536941	1	6	3	4
N	1.397443	1	116.971574	1	-0.285146	1	6	3	2
C	1.428219	1	121.061533	1	-0.160731	1	8	6	3
N	1.353661	1	122.788230	1	1.146092	1	9	8	6
N	1.391608	1	119.523972	1	-172.801788	1	9	8	6
H	0.992877	1	117.117524	1	-29.660571	1	11	9	8
H	0.994146	1	116.443857	1	-171.961795	1	11	9	8
H	1.000010	1	118.975153	1	7.132850	1	8	9	11
H	1.098673	1	125.019632	1	179.951688	1	5	4	3
H	0.990142	1	126.487935	1	0.491123	1	1	2	10
C	1.500000	1	120.335792	1	0.532075	1	7	6	3
H	1.092961	1	114.362647	1	-1.994947	1	17	7	3
H	1.091914	1	115.572006	1	110.910493	1	17	6	3
H	1.091932	1	116.973143	1	-114.366043	1	17	6	3
N	3.842854	1	154.753070	1	-177.300589	1	17	7	3
N	1.098006	1	178.689838	1	25.542596	1	21	17	3
O	0.000000	0	0.000000	0	0.000000	0	0	0	0

**MOPAC 5.0 transition state optimization (MDO6G reaction)
via Bartels's method**

PM3 PRECISE T=1000M MMOK NOXYZ NOINTER NLLSQ XYZ CHARGE=+1
BARTELS'S METHOD - METHYLDIAZONIUM ION ATTACKING O6 OF GUANINE

N	0.000000	0	0.000000	0	0.000000	0	0	0	0
C	1.397324	1	0.000000	0	0.000000	0	1	0	0
C	1.405819	1	106.233241	1	0.000000	0	2	1	0
N	1.405478	1	109.009908	1	0.040483	1	3	2	1
C	1.342877	1	107.393922	1	359.920237	1	4	3	2
C	1.428263	1	120.852808	1	180.917220	1	3	2	1
O	1.252073	1	126.767618	1	177.250557	1	6	3	2
N	1.415952	1	114.933479	1	358.116663	1	6	3	2
C	1.430499	1	120.764951	1	1.661004	1	8	6	3
N	1.346953	1	123.951227	1	0.195589	1	9	8	6
N	1.400245	1	119.565355	1	185.658259	1	9	8	6
H	0.992735	1	115.817904	1	327.437504	1	11	9	8
H	0.994922	1	115.646841	1	188.923978	1	11	9	8
H	0.996763	1	119.785909	1	182.289269	1	8	6	3
H	1.095682	1	124.884961	1	180.043200	1	5	4	3
H	0.987972	1	126.391575	1	179.591960	1	1	2	3
C	2.050696	1	118.278376	1	0.796991	1	7	6	3
H	1.118508	1	95.887694	1	2.305395	1	17	7	6
H	1.089898	1	69.254962	1	122.870648	1	17	7	6
H	1.089598	1	90.917016	1	244.771452	1	17	7	6
N	1.904851	1	157.664529	1	135.099683	1	17	7	6
N	1.099699	1	169.680076	1	354.453007	1	21	17	7
O	0.000000	0	0.000000	0	0.000000	0	0	0	0

**MOPAC 6.0 eigenvector-following transition state
optimization (MDO6G reaction) and calculation of atomic
charges fitted to an electrostatic potential (ESP)**

PM3 PRECISE T=1000M MMOK NOXYZ NOINTER TS XYZ ESP CHARGE=+1
TS (MOPAC6) OPTIMIZATION OF MDO6G AND CALCULATION OF ESP CHARGES

N	0.0000000	0	0.000000	0	0.000000	0	0	0	0
C	1.3952590	1	0.000000	0	0.000000	0	1	0	0
C	1.4060321	1	106.339165	1	0.000000	0	2	1	0
N	1.4064759	1	108.986210	1	0.050994	1	3	2	1
C	1.3441040	1	107.315610	1	0.000000	1	4	3	2
C	1.4299581	1	120.647774	1	-179.939080	1	3	2	1
O	1.2519556	1	128.052736	1	179.650642	1	6	3	2
N	1.4151223	1	114.877089	1	-0.118947	1	6	3	2
C	1.4290515	1	121.107052	1	-0.568349	1	8	6	3
N	1.3474238	1	123.739873	1	1.663267	1	9	8	6
N	1.4009020	1	118.805904	1	-172.497452	1	9	8	6
H	0.9938629	1	115.819147	1	-33.739170	1	11	9	8
H	0.9946717	1	115.537882	1	-171.055611	1	11	9	8
H	0.9984673	1	119.562495	1	-179.557843	1	8	6	3
H	1.0964005	1	125.039510	1	179.983808	1	5	4	3
H	0.9890847	1	126.515833	1	-179.849060	1	1	2	3
C	2.0639915	1	118.440376	1	0.643616	1	7	6	3
H	1.0884971	1	84.660906	1	-121.001425	1	17	7	6
H	1.1171561	1	90.005909	1	-0.495533	1	17	7	6
H	1.0885303	1	84.541335	1	120.041393	1	17	7	6
N	1.9092423	1	176.286925	1	178.185717	1	17	7	6
N	1.0995495	1	178.488095	1	121.837389	1	21	17	18

**MOPAC 6.0 eigenvector-following geometry optimization
(guanine) and calculation of atomic charges fitted to an
electrostatic potential (ESP)**

PM3 PRECISE T=1000M MMOK NOXYZ NOINTER EF ESP
EF (MOPAC6) MINIMIZATION OF GUANINE AND CALCULATION OF ESP CHARGES

N	0.0000000	0	0.000000	0	0.000000	0	0	0	0
C	1.3940441	1	0.000000	0	0.000000	0	1	0	0
C	1.4048539	1	106.675425	1	0.000000	0	2	1	0
N	1.3960654	1	108.565265	1	0.064345	1	3	2	1
C	1.3437624	1	108.010493	1	-0.028500	1	4	3	2
C	1.4486724	1	120.198170	1	179.814350	1	3	2	1
O	1.2158488	1	131.693689	1	-0.302785	1	6	3	4
N	1.4536717	1	112.808102	1	0.203380	1	6	3	2
C	1.4130435	1	122.579052	1	-0.776999	1	8	6	3
N	1.3390596	1	123.909049	1	1.641595	1	9	8	6
N	1.4225865	1	118.454581	1	-173.512074	1	9	8	6
H	0.9960703	1	113.069499	1	-42.884146	1	11	9	8
H	0.9960022	1	113.446545	1	-171.744218	1	11	9	8
H	0.9982476	1	118.540933	1	4.277275	1	8	9	11
H	1.0931992	1	125.039648	1	-179.992369	1	5	4	3
H	0.9872254	1	126.452410	1	0.545143	1	1	2	10

Gaussian 90 MP2/6-31G* energy minimization of formamide

```
#n mp2=(semidirect,maxdisk=8000000)/6-31G* scf=direct opt
```

```
formamide - mp2/6-31G* opt
```

```
O 1
H
N 1 NH1
C 2 CN 1 CNH
H 3 CH 2 HCN 1 HCNH
H 2 NH2 3 HNC 4 HNCH
O 3 CO 2 OCN 5 OCNH
```

```
NH1 1.0
CN 1.4
CH 1.0
NH2 1.0
CO 1.4
CNH 120.0
HCN 120.0
HNC 120.0
OCN 120.0
HCNH 180.0
HNCH 0.0
OCNH 180.0
```


**Gaussian 90 HF/6-31G* transition state optimization for
the methylation of formamide by the methyldiazonium ion**

```
#n hf/6-31G* scf=direct opt=(ts,calcfc)
```

```
transition state of [mdz -- form]+ at hf/6-31g*
```

```
1 1
H
N 1 NH
C 2 NC 1 121.3
H 3 1.108 2 118.5 1 180.0
H 2 0.994 3 121.6 4 0.0
O 3 CO 2 118.5 5 180.0
C 6 OC 3 COC 2 180.0
H 7 1.097 6 HCO1 3 -4.0
H 7 1.097 6 HCO2 3 116.0
H 7 1.097 6 HCO2 3 -124.0
X 7 1.0 6 90.0 3 -4.0
X 11 1.0 7 90.0 6 180.0
N 7 CN 11 87.8 12 0.0
X 13 1.0 7 90.0 11 0.0
X 14 1.0 13 90.0 7 180.0
N 13 NN 14 89.5 15 0.0
```

```
NH 0.996
NC 1.354
CO 1.259
OC 2.081
CN 1.931
NN 1.103
COC 136.5
HCO1 88.8
HCO2 85.4
```

DGauss energy minimization of formamide

```
OPT VWN BP
BASIS=DZVP2 AUX=A2
::
CHARGE 0
MULTIP 1
INTACC medium
DYNACC
XCGRID medium
NITMAX 100
CVSCF medium
NFUNCT 60
CVGRAD medium
MICRO 1
NPOINT 2

MULTMOM mass
MULLIKEN
MAYER
PLOTGRID 4.0 4.0 4.0 30 30 30

GEOMETRY
ANGSTROM
COORD
6 3.32318377 1.77236474 0.00000000
8 4.82318354 1.77236474 0.00000000
1 2.81318378 2.65571070 0.00000000
7 2.56318378 0.45600614 0.00000000
1 1.56318378 0.45600614 0.00000000
1 3.06318378 -0.41001925 0.00000000
END
CARTES
ENDGEO

ENDINP
```

**DGauss transition state optimization for the methylation
of formamide by the methyldiazonium ion**

```
TRANS VWN BP
BASIS=DZVP2 AUX=A2
::
CHARGE 1
MULTIP 1
INTACC medium
DYNACC
XCGRID medium
NITMAX 120
CVSCF medium
NFUNCT 60
CVGRAD medium
MICRO 1
NPOINT 2

MULTMOM mass
MULLIKEN
MAYER
PLOTGRID 4.0 4.0 4.0 30 30 30

GEOMETRY
ANGSTROM
COORD
8 0.71567374 1.18224955 -2.59292459
6 0.68579072 1.93957555 -3.58974266
1 0.64458513 3.04823995 -3.49266458
7 0.70385087 1.47417796 -4.83734751
1 0.67425489 2.10999775 -5.63893461
1 0.74622422 0.46405277 -4.99559069
7 0.73827964 1.72714889 1.13897645
7 0.76319724 1.81041980 2.25132179
6 0.72358894 1.58199704 -0.54889864
1 1.80826712 1.51058030 -0.69623971
1 0.21330537 2.52964854 -0.76267403
1 0.14367409 0.65452379 -0.59849423
END
CARTES
ENDGEO

HESSIAN

ENDINP
```

References

1. Stewart, J.J.P. (1989) MOPAC Manual (Fifth Edition) A General Molecular Orbital Package, Air Force Systems Command, United States Air Force.
2. Stewart, J.J.P. (1990) MOPAC Manual (Sixth Edition) A General Molecular Orbital Package, Air Force Systems Command, United States Air Force.
3. *Gaussian 90*, M.J. Frisch, M. Head-Gordon, G.W. Trucks, J.B. Foresman, H.B. Schlegel, K. Raghavachari, M.A. Robb, J.S. Binkley, C. Gonzalez, D.J. Defrees, D.J. Fox, R.A. Whiteside, R. Seeger, C.F. Melius, J. Baker, R.L. Martin, L.R. Kahn, J.J.P. Stewart, S. Topiol, and J.A. Pople, Gaussian, Inc., Pittsburgh, PA 1990.
4. *UniChem Chemistry Codes - SG-5151 1.0* (1990) Cray Research, Inc.

Key References

1. Lawley, P.D. (1989) *Mutation Research*, 213, 3-25.
2. Margison, G.P. and O'Connor, P.J. (1979) in *Chemical Carcinogens and DNA. Volume 1*, CRC Press, pp.111-159.
3. Singer, B. and Grunberger, D. (1983) *Molecular Biology of Mutagens and Carcinogens*, New York, Plenum Press.
4. Richardson, F.C. and Richardson, K.K. (1990) *Mutation Research*, 233, 127-138.
5. Singer, B. and Kusmierk, J.T. (1982) *Annual Review of Biochemistry*, 51, 655-693.
6. Ford, G.P. and Scribner, J.D. (1983) *Journal of the American Chemical Society*, 105, 349-354.
7. Frecer, V. and Miertus, S. (1989) *Neoplasma*, 36, 257-272.
8. Scribner, J.D. and Ford, G.P.,(1982) *Cancer Letters*, 16, 51-56.
9. Ford, G.P. and Smith, C.T. (1987) *Journal of the Chemical Society, Chemical Communications*, 44-45.
10. Loveless, A. (1969) *Nature*, 223, 206-207.
11. Horsfall, M.J., Gordon, A.J.E., Burns, P.A., Zielenska, M., van der Vliet, G.M.E., and Glickman. B.W. (1990) *Environmental and Molecular Mutagenesis*, 15, (1990) 107-122.
12. Parthasarathy, R. and Fridey, S.M. (1986) *Carcinogenesis*, 7, 221-227.
13. Leonard, G.A., Thomson, J. Watson, W.P., and Brown, T. (1990) *Proceedings of the National Academy of Sciences*, 87, 9573-9576.
14. Ludlum, D.B. (1970) *Journal of Biological Chemistry*, 245, 477-482.

15. Schelgel, H.B. (1987) in Ab Initio Methods in Quantum Chemistry - I, Lawley, K.P., ed., Wiley & Sons, pp. 249-285.
16. Stewart, J.J.P. (1990) *Journal of Computer-Aided Molecular Design*, 4, 1-105.
17. Stewart, J.J.P. (1988) MOPAC Manual (Fifth Edition) A General Molecular Orbital Package, Air Force Systems Command, United States Air Force.
18. Stewart, J.J.P. (1990) MOPAC Manual (Sixth Edition) A General Molecular Orbital Package, Air Force Systems Command, United States Air Force.
19. Parr, R.G. and Yang, W. (1989) Density-Functional Theory of Atoms and Molecules, Oxford University Press, New York.
20. Unichem Chemistry Codes - SG-5151 1.0 (1991) Cray Research, Inc.
21. Andzelm, J. (1991) in Density Functional Methods in Chemistry, Labanowski, J.K. and Andzelm, J.W., eds., Springer-Verlag, New York.
22. *Gaussian 90*, M.J. Frisch, M. Head-Gordon, G.W. Trucks, J.B. Foresman, H.B. Schlegel, K. Raghavachari, M.A. Robb, J.S. Binkley, C. Gonzalez, D.J. Defrees, D.J. Fox, R.A. Whiteside, R. Seeger, C.F. Melius, J. Baker, R.L. Martin, L.R. Kahn, J.J.P. Stewart, S. Topiol, and J.A. Pople, Gaussian, Inc., Pittsburgh PA, 1990.